2014

Impacts of Complexity and Timing of Communication Interruptions on Visual Detection Tasks

Sally Stader
University of Central Florida

IMPACTS OF COMPLEXITY AND TIMING OF COMMUNICATION INTERRUPTIONS ON VISUAL DETECTION TASKS

by

SALLY A. STADER
M.S. University of Central Florida, 2005

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2014

Major Professor: Mustapha Mouloua

ABSTRACT

Auditory preemption theory suggests two competing assumptions for the attention-capturing and performance-altering properties of auditory tasks. In onset preemption, attention is immediately diverted to the auditory channel. Strategic preemption involves a decision process in which the operator maintains focus on more complex auditory messages. The limitation in this process is that the human auditory, or echoic, memory store has a limit of 2 to 5 seconds, after which the message must be processed or it decays. In contrast, multiple resource theory suggests that visual and auditory tasks may be efficiently time-shared because two different pools of cognitive resources are used. Previous research regarding these competing assumptions has been limited and equivocal. Thus, the current research focused on systematically examining the effects of complexity and timing of communication interruptions on visual detection tasks. It was hypothesized that both timing and complexity levels would impact detection performance in a multi-task environment.

Study 1 evaluated the impact of complexity and timing of communications occurring before malfunctions in an ongoing visual detection task. Twenty-four participants were required to complete each of the eight timing blocks that included simple or complex communications occurring simultaneously, and at 2, 5, or 8 seconds before detection events. For simple communications, participants repeated three pre-recorded words. However, for complex communications, they generated three words beginning with the same last letter of a word prompt. Results indicated that complex communications at two seconds or less occurring before a visual detection event significantly impacted response time with a 1.3 to 1.6 second delay compared to all the other timings. Detection accuracy for complex communication tasks under the simultaneous condition was significantly degraded compared to simple communications at five seconds or more prior to the task. This resulted in a 20% decline in detection accuracy. Additionally, participants' workload ratings for complex communications were significantly higher than simple communications.

Study 2 examined the timing of communications occurring at the corresponding seconds after the visual detection event. Twenty-four participants were randomly assigned to the communication complexity and timing blocks as in study 1. The results did not find significant performance effects of timing or complexity of communications on detection performance. However the workload ratings for the 2 and 5 second complex communication presentations were higher compared to the same simple communication conditions.

Overall, these findings support the strategic preemption assumption for well-defined, complex communications. The onset preemption assumption for simple communications was not supported. These results also suggest that the boundaries of the multiple resource theory assumption may exist up to the limits of the echoic memory store. Figures of merit for task performance under the varying levels of timing and complexity are presented. Several theoretical and practical implications are discussed.

For Jack, Indy, Edo, and airplanes.

ACKNOWLEDGMENTS

While the mantle of the dissertation is carried by a single author, the work is worn by many. There are many people who I wish to acknowledge in this work. I stand in awe of the collective wisdom of my dissertation committee. To my advisor, Dr. Mustapha Mouloua, I cannot thank you enough for your patience and persistence in seeing me through. I am truly honored to have been your student and to now carry on the traditions and practices of human factors. To Dr. Peter Hancock, your keen questioning, amazing depths of knowledge and steadfast determination to seek truth have been a source of great inspiration. Thank you for sharing all of these during this dissertation process. To Dr. Mark Neider, thank you for the insightful ideas, encouragements, review of the manuscript, and guidance throughout the course of this dissertation. I have sincerely appreciated your enthusiasm for research and willingness to consult during this project. To Dr. Peter Kincaid, yours was my first graduate class at UCF and inspired me to continue on to the doctorate. Words cannot fully express my deep gratitude for your guidance throughout these graduate school years. I also wish to thank Dr. Richard Gilson, my first advisor. Your vitality, courage, optimism, and contributions to the field of human factors have inspired me and so many others.

To fellow aviator and co-researcher, Jennifer Leavens, this research would not have been finished if not for your friendship and support. Thank you for being available at a moment's notice throughout this project. Thank you to the members of the UCF Transportation Research Group lab and to all of my fellow graduate students over these years. Thanks to my friend, Yaela Dahan Marks. The moral support and encouragement have been very much appreciated.

To my wonderful husband, Jack McCloy, the commitment you have shown to me throughout these years of graduate school has been truly remarkable. Without your encouragement, moral support, patience, and technical expertise, this research would never have happened and this document would not exist.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Statement of the Problem

Each year the National Transportation Safety Board issues their list of "Most Wanted" critical changes that are needed to reduce transportation accidents and save lives (NTSB, 2014). For the past two years the list has included the elimination of distractions in transportation systems because of various safety concerns.

One known distraction related to driving tasks involves interruptions from conversations such as those on a cell phone (Horrey, & Wickens, 2006) or from other sources such as listening to books on tape or to music, voice dialing and searching for information by voice, and listening to and orally replying to questions (Angell, Auflick, Austria, Kochhar, Tijerina, Biever, Diptiman, Hogsett, & Kiger, 2006).

Previous research regarding the deleterious effects of cell phone use on driving reported that cell phone conversations significantly disrupt driving tasks in terms of increased probability of missing a traffic signal, increased time to respond to a signal, increased error in a manual navigation tracking task, slower accelerations to desired speed, slower braking response times, and longer latency in depressing the brake pedal when compared to non-conversation or driving-only conditions (Strayer & Johnston, 2001; Strayer, Drews and Johnston, 2003). The overall effect sizes were generally medium to large (e.g. ES = 0.74 in the 2003 studies) for the impact of conversations on these tasks. Considering the results, Strayer, et al. (2003) concluded that conversations appear to disrupt the driving task due to direction of attention away from the external world and toward an "internal cognitive context" and indicating that a cognitive information processing requirement is the distracting factor.

In the aviation domain, it has also been shown that verbal communications frequently preempt higher priority flight tasks (Damos, 1997; Dismukes, Loukopoulos, & Jobe, 2001), interrupt work flow of the cockpit crew (Loukopoulos, Dismukes, & Barshi, 2001), and are the primary interruption or distraction facing flight crews (Airbus, 2004a). In addition, modern automated aircraft flight decks may not be specifically designed to handle or manage various types of interruptions, including communications (McFarlane & Latorella, 2002).

Incident reports to the voluntary Aviation Safety Reporting System (ASRS) also have indicated that communications may impose significant costs to pilot performance during flights (Connell, 1995) with over 70% of the reports citing problems with transfer of information, and almost 50% of interruptions in flight related to communications, such as from ATC or flight attendants (Damos & Tabachnick, 2001). Accident investigations have confirmed the impact of communication interruptions to flight performance. Failures in communications have been implicated in some of the most catastrophic aviation accidents, with human factors issues related to various forms of interpersonal communications implicated in 70% or more of all accidents in recent years (Sexton & Helmreich, 2000).

In the future National Airspace System under the NextGen (Next Generation) air transport initiative, it is proposed that the air traffic control communication structure will change to one of shared responsibilities for communication and multi-way information exchange between controllers, flight crews, and other entities involved in flight management (e.g., airport gate agents, airlines dispatch and maintenance). While digital data uplink and downlink is already available in commercial aircraft under the proposed, NextGen-inspired, FAA Data Communications (Data Comm) and System-Wide Information Management (SWIM) programs (FAA, 2013a, b), routine communications will be shared among all elements of ATC, the flight deck, airline companies, etc. via digital information technology. These programs are described as supplemental to existing voice communications, and not a total replacement of voice, at least for the foreseeable future.

As routine communications become digital, what will remain for verbal communications for pilots will be the non-routine, time-critical, or emergency situations, and in those cases, responding in the verbal mode may be the difference between a safe flight and a failed one. Such a shift will require a focus on task management skills (Iani & Wickens, 2007) including management of task interruptions (Trafton & Monk, 2007).

Given this future ATC scenario, it is not difficult to imagine aircrew responsibilities becoming increasingly that of supervising and monitoring of automated systems, stepping in as problem-solvers in non-normal situations. In such situations, human are known to experience problems with sustaining attention to tasks, over-trusting or over relying on the system, experiencing imbalances in workload (too much or too little), and mistrust or misuse of systems. Over time, operator skill may be eroded and situation awareness reduced. In short, changes in technology as well as communication interruptions and failures pose significant risks to operator performance in various domains, and research is needed to address these risks.

Together, the studies from the driving and the aviation domains indicate that conversation and communications frequently interrupt operators and pose significant impacts to ongoing tasks across various operational domains. However, little research in these areas of distractions imposed by communications during visual tasks has examined exactly when these interruptions first begin to distract the operator. Understanding these initial moments of distraction may help to inform future design of systems to alert the operator regarding when they are being distracted from their primary task.

While research in these areas has shown that verbally-related auditory interruptions disrupt various vehicle operating tasks, little research has examined exactly what occurs in the moment of the distraction and how those early moments may vary depending on complexity and timing of the interrupting communication.

In addition, there has been little empirical research to examine the impact of communication interruptions to ongoing tasks that consist of events that may be time-critical, such as monitoring an automated system (rather than the environment outside of the system) for malfunctions (i.e. errors). The early detection of system malfunctions can provide an additional safety barrier that can help to reduce human error when operating complex or automated systems (e.g. Sharit, 2005).

An important human factors issue to consider in this type of research is workload assessment. Workload is a multidimensional construct that has eluded an exact definition in human factors research. However, it is generally accepted that workload involves the relationship of an amount of work activity to be completed in a specified time period. Thus, more work in a shorter amount of time would lead to higher workload level. Workload can be measured by metrics that capture the nature of the task in comparison to various levels of that task or to other tasks, and it can be measured subjectively by asking the person to reflect on their perception of their workload. In general, workload is known to be an important consideration when assessing operator performance (Mouloua, Hancock, Jones, & Vincenzi, 2010).

Based on the brief review of literature above, this research will focus on the impact of the complexity and timing of communication interruptions in relation to operator detections of system malfunctions and operator performance of a manual tracking task. Operator subjective workload and responses to communications also will be assessed.

Research Needs Addressed by the Current Study

By examining the impact of communications on ongoing visual monitoring and detection tasks, the current research intends to contribute to two different bases of literature. The first is in regard to multiple resource theory (MRT) versus auditory preemption theory. Several studies related to MRT (e.g. Wickens, 2002) have found that two tasks, one visual and one auditory, may be performed simultaneously with little conflict because they use two different pools of cognitive resources. In contrast, other research (e.g. Wickens & Liu, 1988) has found that auditory tasks can interrupt (preempt) visual and manual tasks—often to a large degree as noted in the earlier review of driving studies—when they are performed simultaneously.

One factor in determining the impact of these interruptions appears to be the type of auditory interruption. That is, a simple warning tone, while distracting momentarily, may impose an interruption at the onset of the tone but little ongoing interruption to a visual task. However, another type of auditory distraction, such as a complex communication, may introduce an increased information processing demand and thus requires more cognitive resources of the operator. What is not known is whether there are definable boundaries based on both complexity and timing of an auditory interruption that may explain exactly when operators may be distracted from their other tasks, especially visual tasks requiring an accurate and quick response.

Therefore, the prior research that has attempted to define when auditory tasks interrupt visual tasks and when they do not is still unresolved and represents a gap in this literature. To further assess this gap, a recent meta-analysis was performed (Lu, Wickens, Prinet, Hutchins, Sarter, & Sebok, 2013). The results were inconclusive and the researchers called for continued research to investigate the moderating variables that may define which theory (i.e. MRT or auditory preemption) operates in which situations. The current study is designed to further the research regarding this question with a specific focus on two levels of complexity of communications as the interrupting auditory tasks and the timing of communications in relation to visual monitoring and detection.

The second intended literature base is regarding research on task interruptions (e.g. Trafton & Monk, 2007) and interruption management (e.g. Latorella, 1999; Iani, & Wickens, 2007). The research in the area of task interruptions has come largely from the human-computer interaction domain with a focus on visual interruptions to mostly ongoing visual tasks in office environments. The research on interruption management is considered an expansion of the task interruption literature and has been largely focused on tasks in the aviation domain.

The research in these two areas define the primary task of interest as the ongoing task (OT), while the interruption is referred to as an interrupting task (IT) denoting its importance as a task itself that cannot be ignored and which requires a response from the operator. One focus of this research has been in determining when these two types of tasks (i.e. the OT and the IT) can be deferred, or when they must be interleaved, for the most successful overall performance. The research described herein is expected to contribute to this literature base by examining how operators respond to a communications interrupting task when required to also perform a primary ongoing visual monitoring and detection task.

Approach for the Current Study

This research was designed to evaluate the impact of the communication task in a multitask environment. More specifically, this study focuses on the impact of the complexity and timing of communication interruptions in relation to operator performance. Performance is measured as detection of and response time to critical system malfunctions and maintaining accuracy on a navigation path. The impact on subjective workload from complexity and timing of communication interruptions is also examined.

To evaluate the two primary communication variables, this research consists of two studies. Each study tests a 2 (communication complexity) x 4 (timing interval) within subjects design. In each experiment, the two levels of communication complexity (i.e. simple or complex) are defined as the level of information processing that is required to formulate a response to a conversation request similar to the Strayer and Johnston studies (2001). The use of an information processing task to represent communications in experimental research was supported in a recent meta-analysis (Horrey & Wickens, 2006).

The intervals for timing are based on the known limitations to the human auditory sensory store (i.e. echoic memory) demonstrated by Treisman (1964). That prior research found that an auditory stimulus can remain in echoic memory for only about 2 to 5 seconds before further processing is required. Therefore, it is expected that the most disruption to a visual task may occur within that time frame. An auditory stimulus presented either before or after 5 seconds may be expected to impose little disruption to visual or manual tasks as the operator may adopt a strategy for managing the interruption. It is expected that the results of this study will add to the two bases of literature as described earlier.

CHAPTER TWO: LITERATURE REVIEW

Conversation and Communication Interruptions

In order to frame the issues regarding interruptions from auditory communications in transportation, it is instructive to examine studies from two domains—driving and aviation. Key research is reviewed below.

Conversations among Drivers and Pedestrians

In an influential study from the driving domain, four experiments were performed to examine the impact of casual, naturalistic hands-free cell phone conversations during automobile driving tasks. Strayer, Drews and Johnston (2003) found that such cell phone conversations significantly disrupt driving. Most related to the current research, their first experiment involved the impact of a conversation on following another car in low traffic versus high traffic density conditions. Results indicated that when engaged in cell phone conversations, participants were slower to accelerate to their desired speed, slower to apply brakes in response to the car in front, and tended to press the brake pedal longer compared to the driving-only task. The overall effect size was reported as medium to large (ES = 0.74). Higher density traffic conditions tended to increase the difference between driving-only and driving-while-conversing conditions.

Considering the results across the four studies, Strayer, et al (2003) argued that cell phone conversations provide a significant distractor to necessary driving tasks—such as braking for an automobile ahead or attending to objects in the environment—due to inattention blindness. That is, conversations appear to disrupt the driving task due to direction of attention away from the external world and toward an “internal cognitive context” (p. 31). They note that there was no manual manipulation of the cell phones during the dual task portions of the study eliminating that factor as a reason for the inattention and supporting earlier studies in this regard (Strayer &

Johnston, 2001). The authors also noted that while the cell phone conversations were designed to be casual and naturalistic, the results may underestimate the impact of other conversations such as business negotiations or emotionally consuming conversations. They stated future research could investigate impacts of different types of conversations on driving tasks.

Finally, the authors discussed that during debriefing, participants showed a “disconnect” between their own self-perception of their driving performance and the objective performance measures that were collected. They summarized by stating, “A consequence of using a cell phone is that it may make drivers insensitive to their own impaired driving behavior” (p. 31).

A meta-analysis by Horrey and Wickens (2006) found that conversations on cell phones and from passengers impose significant costs to driving performance, with the most significant being response times to critical road or driving events, and to a lesser degree to lane-keeping (i.e. tracking) maneuvers. These costs were evident for both hands-free and hand-held devices. Overall, their meta-analysis showed an average response delay of 0.13 seconds. In their summary, Horrey and Wickens (2006) note that tracking and event response time “represent logical precursors to less frequently observed loss-of-control and collision events” (p. 204). They note that other important factors for study include workload and actual accident events.

More recently, Horrey (2011) summarized results of the Driver Workload Metric project (Angell, et al., 2006) in a graphic that depicts combined performance deficits for response times and missed events from in-vehicle tasks and interruptions. Among these in-vehicle tasks were a variety of common conversation-related secondary tasks such as cell phone use, listening to and adjusting radio and CD players, manual and voice dialing, listening to and discussing a book on tape, mentally computing and saying aloud travel distances, and listening to and repeating back route instructions. The graphic indicated a clear relationship between the cognitive complexity of the tasks versus response times to discrete events and missing those events.

Horrey (2011) explained that, on average, the Angell, et al., (2006) data indicates response times to critical events were slowed from between 130 to 210 ms (p. 6). In their technical report, Angell, et al. (2006) note that higher object and event detection (OED) misses were associated with slower response times. And conversely, higher OED percentages of missed events were associated with quicker response times. It should be noted that specific timings of the onset of conversation events was not manipulated in the study.

Research has also examined pedestrians' behavior while using a cell phone and walking across a street (Neider, McCarley, Crowell, Kaczmarski, & Kramer, 2010). Results found that when pedestrians' were talking on a cell phone they were less successful at street crossings compared to listening to music or undistracted crossings, with success defined as completing the crossing within a 30 second time limit. Their initiations of crossings were also delayed by about 1.5 seconds when compared to the other two conditions. A subsequent study (Neider, et al., 2011) found that older adults also showed crossing performance decrements while talking on a cell phone and were comparatively more impaired than younger adults in terms of crossing initiation delays and timing out in the crossing.

Communication as an Aviation Task

Pilots when flying must follow much stricter rules than drivers when it comes to conversation. Conversation has an informal connotation in the driving domain and implies informality as well as unnecessary distractions to the ongoing tasks. However, in the aviation domain the act of sharing information is a task in itself. Thus conversations are typically vital communication acts for operators in aviation. This distinction underscores the emphasis of the current research.

At this point in aviation history, it is important to study communications for two reasons. First, technology is significantly changing the flight environment. Current technology allows routine communications to be uploaded in a visual, digital format, historically called DataLink (recently renamed DataComm; FAA, 2013). Such routine communications may consist of flight clearances, changes to flight plans at later phases of a flight, or other information or instructions that are not immediate in nature.

However, a concern with digital communication is that what remains for radio broadcast via voice are often non-routine or time-critical messages which can be unpredictable and stressful (Morrow, Rodvold & Lee, 1994; McGann, et al. 2009). Non-routine messages in the flight environment usually require several communication turns between the speakers, each turn containing several speech acts, particularly when a message is misunderstood (Prinzo & Britton, 1993). In addition, a study by Harvey, Reynolds, Pacley, Koubek, and Rehmann (2002) found that while DataLink can *decrease ATC-to-flight-crew* voice messages, the DataLink messages from ATC actually *increase within-crew* voice communication because crew members must read the information transmitted via DataLink and discuss it in order to make decisions regarding the flight. This potential increase—not decrease—in the voice communication load in the cockpit requires a better understanding of how voice communication impact other tasks that pilots perform during a flight.

Second, communication errors can impose significant costs to pilot performance during flights, and ultimately to the safety of flights. For example, Connell (1995) studied reports to the Aviation Safety Reporting System (ASRS), a voluntary aviation incident reporting system, and found that in the first five years of the ASRS existence (from 1976 to 1981), “over 70% of the reports submitted noted problems in the transfer of information” (p. 20). Damos and Tabachnick (2001) studied reports submitted to the ASRS between 1991 and 1998 regarding interruptions in the cockpit and impacts on flight crew performance. They found that almost 50% of the interruptions were related to communication (from ATC or flight attendants). In addition, 55.6% of the time, communications from ATC interrupted both pilots of two-person crews, and 80% of the time more than one pilot was interrupted in three-person crews (p. 20).

Accident investigations have confirmed the impact of communication interruptions to flight performance that have been voluntarily reported to ASRS. Failures in communication have been implicated in some of the most catastrophic aviation accidents (for example, see Cushing, 1994; Kanki & Palmer, 1993; Helmreich, 1997; Krivonos, 2007). Sexton and Helmreich (2000) reported, “Human factors issues related to interpersonal communication have been implicated in approximately 70% to 80% of all accidents over the past 20 years” (p. 63). Krivonos (2007) cited a Flight Safety Information (2004) report that found, “between 1976 and 2000, more than 1100 passengers and crew lost their lives in accidents in which…language played a contributing role” (p. 4). Communication failures reduce team coordination and decision-making (Serfaty, Entin, & Volpe, 1993; Orasanu, Martin & Davison, 2002), which in turn contributes to poor management of all flight tasks (Iani & Wickens, 2007). All of these studies together indicate that technology changes and communication interruptions and failures pose significant risks to aviation safety.

Communication Complexity, Timing, and Types of Errors

To date, various characteristics of aviation communications and their impact on pilot performance have been studied. These characteristics include issues of complexity (i.e. message length and format, improving pilot recall for lengthy messages after errors); issues of timing (i.e. communication by phases of flight, timing between communication acts); and classifications of errors by types of communication acts.

One of the most robust findings from aviation communications research is that longer messages (i.e. five or more discrete instructions per single message) impose demands on a pilot's limited working memory and require more requests for clarification than shorter messages (i.e. four or fewer discrete instructions per communication; Morrow & Rodvold, 1993; Morrow Rodvold, & Lee, 1994; Burki-Cohen, 1995; Prinzo & Morrow, 2002). This finding holds regardless of pilot experience and despite attempts to shorten or "chunk" the format of instructions that contain numbers (i.e. saying numbers in groups such as "thirty-four-hundred", versus a sequential number format such as "three-thousand-four-hundred"; Prinzo & Morrow, 2002). Other research has found that when a communication misunderstanding occurs, restating the message improves pilot recall and the format of the restatement appears to matter little, (Burki-Cohen, 1995).

The issue of message complexity is the first variable of interest in the current research. Based on the summary above, the existing aviation research has considered message complexity only in terms of the number of speech acts per message that have been found to produce misunderstandings. Little existing research has addressed the complexity of communication other than message length, or in relation to performance of other ongoing tasks in the cockpit.

Several studies have addressed timing at the macro-level by examining the characteristics of communication between air traffic controllers and pilots in various phases of flight including tower to ground (Burki-Cohen, 1996), tower to air and en route (Cardosi, 1993, 1996), approach control (Prinzo, 1996) and terminal radar approach control (Cardosi, Brett, & Han, 1996).

Cardosi, Falzarano, and Han (1998) synthesized these studies and offered several practical recommendations. For example, controllers should speak slowly and distinctly, keep instructions short with no more than four instructions per transmission, actively listen to pilot readbacks and check for accuracy (i.e. "hearback" pilot messages), point out when call signs are similar among aircraft on the frequency, and avoid issuing strings of instructions to pilots. In addition, pilots should respond to controller instructions with full readback and should always give their call signs with readbacks so that controllers can assure the messages were delivered accurately.

In terms of micro-level message timing issues, Morrow and Rodvold (1993) found an interaction between ATC message length and timing such that shorter time between messages of longer length increased the potential for misunderstandings, though requests for readback reduced the miscommunication. This is one of only a few studies that have addressed the issue of the timing of messages at this level of analysis.

In terms of types of errors related to a macro-level timing of communication acts, Cardosi, Falzarano, & Han (1998) analyzed 386 ASRS reports and classified errors into 3 types: Readback/hearback errors which would occur after a requested communication (47%), no pilot readback (25%), and hearback errors type II (i.e. when ATC fails to notice their own errors or fails to correct critical pilot errors in statement of intent, 18%).

The issue of message timing is the second variable of interest in the current research. Based on the review above, existing research has addressed timing of communication mostly from the macro-level (e.g. in phases of flight). In part, the lack of studies on communication timing may be due to the fact that many studies of communication in aviation are descriptive, such as the studies that have examined reports to the ASRS database, rather than experimental. Manipulating the timing of communication and measuring its effects is inherently an empirical problem. However, the micro-level issues of pilots repeating their call signs (Monan, 1983) and the controller "hearback" problem (Monan, 1988) have emerged from this line of research and are frequently mentioned as important factors in maintaining safety and reducing errors in the flight environment.

Task Interruptions

Task interruptions, as well as multitasking, have received much attention in research and in popular media in recent years. Trafton and Monk (2007) cite an estimate given by Spira & Feintuch (2005) that interruptions cost around $588 billion dollars a year in lost worker productivity. In addition, most of the research in this area has focused on computer-related interruptions and have come from the human-computer interaction domain. Thus much early research focused on office environments and computer applications with many studies investigating tasks that are visual in nature. However, research on interruptions in other domains such as aviation, medicine, and driving has expanded in the past 10 to 15 years, according to Trafton and Monk (2007). Primary characteristics of interruptions that appear to affect performance of primary tasks include complexity, duration, timing and frequency.

Trafton and Monk (2007) noted that the theory they proposed, called memory for goals theory, was developed using a classic psychological laboratory task (i.e. the Tower of Hanoi task). They stated that different real-world tasks, such as tasks that rely on communications which do not have visual environment cues, might not conform to the principles or predictions of their theory. They suggested that these are questions for empirical research. The research proposed for the current studies have testing of one element of this theory as a primary goal.

While relatively few task analyses exist in the interruptions domain, naturalistic observations across several domains led Trafton and Monk (2007) to develop a time line that depicts task disruptions. This time line starts with an ongoing, visual task interrupted by a secondary task which causes the primary task to be suspended for a time while the secondary task is attended or completed. A representation of this time line is presented in Figure 1.

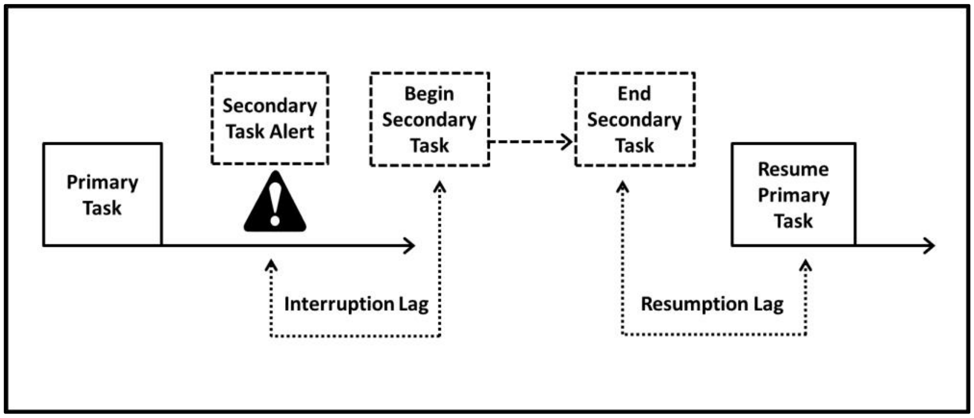

Figure 1. Representation of the task interruption time line.
Source: Trafton & Monk, 2007.

Research in the area has found two lag periods in this time line. The first has been called an interruption lag and it occurs at the alert of the secondary task and before it starts. During this time the operator is mentally shifting toward dealing with the secondary task and may use environmental cues to enable recall of the primary visual task when it is resumed.

The second lag has been termed the resumption lag and it occurs after the secondary task is accomplished and before the primary task resumes. During this time, the operator is reorienting to the primary task and what was happening within that task prior to the interruption.

This model assumes that executive control and task switching are important to the process, as well as all aspects of memory including short-term working memory, long-term working memory and memory stores, retrospective memory for what was accomplished in the primary task, and prospective memory for next steps in the primary task.

According to this model, Trafton and Monk (2007) discussed five aspects of the cognitive system that are involved in interruptions. They included the following.

1) Executive control is important for all tasks.

2) After the secondary task is completed, the person must remember the primary task and what comes next in that task.

3) Environmental cues may or may not be available to aid in remembering what comes next in the primary task.

4) The primary and secondary task may or may not be related.

5) Depending on the interruption, environmental cues may not have been considered, and different preparatory processes may occur (Trafton & Monk, 2007, p. 114).

Perhaps the most robust finding regarding the disruptiveness of an interruption to a primary visual task has to do with the duration of that interruption. In a series of experiments, Monk, Trafton, and Boehm-Davis (2008) found that longer interruptions are more disruptive, as measured by the amount of time needed to resume the primary task (i.e. the duration of the resumption lag). Shorter interruption durations of 3, 8, and 13 seconds had much less of an impact on the duration of the resumption lag than longer interruption durations of 23, 38, and 58 seconds. Resumption lag data has been found to follow an exponential curve for longer interruptions. This finding indicated that future research should consider, or laboratory experiments should hold constant, the duration of the interruption in order to examine effects on the primary task. Inherent in this finding is that the timing of interruptions should also be considered because longer interruptions (i.e. more than 13 seconds) appear to exponentially impact the resumption of the primary task and may potentially impact the operator's detecting of important changes that are occurring in the primary task while the interrupting task is attended.

Much of the research and theories of interruptions to date operationalize “interruption” as the impact to the resumption lag for the primary task. Therefore a good deal is known about the impacts of interruptions at the “far end” end of the interruption/resumption process, especially regarding single-modality (typically visual) interruptions to single visual primary tasks. This far-end process involves the short-term working memory to long-term working memory storage links that allow the person to recall and resume the primary task after a disruption.

Based on this review of the literature, the primary tasks typically studied have been continuously displayed visual tasks, and interrupting tasks have also been visual. In their summary, Trafton and Monk (2007) stated that while there are several differences among interruption theories, they all are based on memory processes. They stated that none of the theories have an emphasis on perception, action, or other cognitive functions.

What is notable from this review is that very little of the theoretical research has focused on the “near end” of the interruption process, that is the cognitive processes involved prior to or during the interruption lag for the secondary task. Those processes would rely on earlier elements of human information processing and memory including attention and sensory memory stores. Also the literature is lacking theoretical studies that examine both secondary tasks and primary tasks that are time-critical, that have one or other task that is discrete or is continuous with periodic discrete events, or that include both visual and auditory tasks. And as stated earlier, another aspect involves the processes involved when the secondary interrupter task is auditory, such as communications, where there are no persistent reminders of what needs to be accomplished.

Interruptions as Operational Tasks

In the driving domain, the primary task is well understood—navigating the vehicle and arriving at the destination safely. In this context, virtually any other stimulus is an interruption, and these interruptions come from many sources. In the quest to make driving safer and more enjoyable, newer automobile designs include integration of portable devices, as well as technology within the vehicle, that provide comfort to the driver, assist with navigation to the destination, and even aid in maintenance of safe distances from other vehicles. These technologies include both visual and auditory signals to the driver that can become distracting, despite important safety information that may be conveyed.

In light of these desired innovations, human factors researchers warn of the increase in driver distractions from in-vehicle technologies that provide benefits, but also impose costs (Horrey, 2011). The costs may include visual and auditory interruptions, both distractions (i.e. passenger conversations, voice dialing to request information, listening and tuning radios or CD players) and secondary tasks that may help to aid navigation and safety (i.e. a talking GPS, warnings from lane-keeping systems, visual monitoring of the in-vehicle devices). As reviewed earlier, much recent research regarding distractions and driving have focused on the use of cell phones, and many of the variables of interest have been lane-keeping (e.g., Horrey & Wickens, 2006) and response times to a lead vehicle (e.g., Levy & Pashler, 2008).

In aviation, maintaining a safe flight involves ongoing but predicable procedures and rules. The tasks involved with the ongoing maintenance of safe flight are largely visual or visual-motor in nature. Such tasks may include monitoring engine characteristics such as pressure and temperature, evaluating fuel levels and adjusting balance of fuel in the tanks or time left in the flight based on fuel remaining, and visually assessing the path of the aircraft and making manual inputs to maintain the desired, projected path. Interruptions to these ongoing tasks (OT) in flight are known to impose certain costs in terms of increased workload, increased time to perform tasks, and reduced accuracy of task performance (Lu, Wickens, et al., 2013) as well as reduced situation awareness (Airbus, 2004b).

When interruptions are evaluated as not important to the current task or to tasks in the foreseeable future, they can be delayed to a later, more convenient time, or they can be ignored. However, interruptions that are either critical to the current task or to characteristics of the flight in the foreseeable future become tasks themselves and must be interleaved into the current, ongoing tasks. Interruptions that are seen as tasks (and not nuisance interruptions) are referred to as interrupting tasks, or IT (Iani & Wickens, 2007). Various studies have examined modalities of IT in relation to modalities of the OT (Lu, Wickens, et al, 2013). Modality refers to the sensory processes that are involved in human information processing in completion of tasks. Modalities are auditory, visual, and tactile senses and their processes.

Distractions and interruptions have specific and negative impacts on human performance. In their book, Loukopoulos, Dismukes, and Barshi (2009) present studies of the impact of interruptions on multitasking in the aviation environment. For example, observations of air crews during flights have found that interruptions during various phases and procedures in the flight environment are common (Loukopoulos, Dismukes, & Barshi, 2003). Of particular relevance to the current study, the authors explain that crews are frequently interrupted during preflight and taxi procedures by radio communication. They state that, "The timing of the interruptions and the nature of the response required is largely unpredictable, which means that the crew has little chance to plan in advance how to interleave the interrupting activity into the ongoing flow of tasks (p. 3)."

Interruptions from Communications

Operational context features have been identified as factors that may contribute to aircraft accidents (Dismukes, Berman & Loukopoulos 2007; Orasanu et al. 2002). Among these, critical context features in aviation incidents and accidents include time pressure and disruptions to tasks (such as communications disruptions, Mosier et al. 2010).

Mosier and colleagues (2010) asked a group of airline pilots to complete ratings of the impact on six human-automation interaction (HAI) consequences after reading scenarios based on previous aircraft incident reports. The scenarios were specifically designed to examine the interplay among features of automation, task and context. The researchers found that the pilots perceived the context factor of communication disruptions (e.g., ATC issuing a traffic advisory while the automation monitoring pilot was engaged in a late clearance change to the flight management system) as significantly contributing to increases in three of the six HAI consequences (i.e., workload, effort involved in task management, and potential for automation-related error). The researchers proposed that such context factors may exacerbate the effects of HAI consequences imposed by certain task and automation features.

Multitasking in Operational Environments

In the flight environment, pilots use a rule of thumb regarding prioritization and management of flight tasks often termed the "aviate, navigate, communicate" rule. Aviate is the cardinal rule for a very good reason: the most important task by far is to keep the aircraft upright and stable in flight. In close second is the navigate task, or to process various information—such as instruments, charts, and the view outside the cockpit—in order to identify "both hazardous objects to be avoided (other aircraft, terrain) and objects to seek (e.g. a runway)" (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003, p. 361). Both tasks are nearly entirely dependent on the visual system which makes vision the primary sensory modality in flight environments.

Given these two critical visual tasks (keep the airplane flying and look for objects), it is little wonder that communicate—talking to others both within and outside the cockpit, often called "radio work"—falls to third place, or what Wickens, et al. (2003) term a "side task" in the flight environment. As Wickens, et al. (2003) note, it is not that this task is unimportant. It is that the lower priority "side tasks" should give way to the higher priority tasks when there is conflict for limited attentional resources.

However, ignoring or minimizing the communication task is not without its hazards. For example, Orasanu, Martin and Davidson (2002) presented a summary table of the distribution of types of errors that were found in a study of 37 airline accidents analyzed by the NTSB from the period of 1978 through 1990 (NTSB, 1994). This error classification listed the category of communication as sixth among eight primary error categories, accounting for 4.3% of total errors in crew performance in the 37 accidents. While communication may be considered a lower priority task, multitasking that includes communications is important to the overall goal of a safe flight. Similarly multitasking has become much more prevalent in the driving environment.

Tasks in Automated Systems

Rasmussen (1983) presented a model for describing three types of human performance that are required when performing tasks and their relations to modes of error that are found in modern technical systems. Rasmussen (1983) identified the three types of performance as skill-based, rule-based and knowledge-based behaviors.

The skill-based behavior is described as sensory-motor performance that is mostly outside of conscious control and is performed automatically and smoothly. An example would be control of an aircraft using a yoke or a stick.

In rule-based behavior, a routine or procedure during a task is performed based on some stored or communicated rule regarding completion of a task. The rule-based behavior generally involves carrying out some procedures or acts in a structured manner toward attainment of a goal that is relevant to the task at hand. An example would be using certain fuel level rules to determine when to move the fuel selector to transfer fuel between tanks in an aircraft.

Finally knowledge-based behavior occurs more often in unfamiliar situations or tasks for which no known rules are available from other sources or from the person's experience. According to Rasmussen, in this type of behavior, a person is thought to develop a "mental model" as well plans that govern future actions toward completion of goals given the set of circumstances encountered. An example would be monitoring an automated system and making decisions regarding the stated of the system and actions to take when errors occur.

The tasks in the simulated system used for this study were classified using Rasmussen's typology and are further described in Chapter 3. Such a typology may prove useful for analyzing the results from this research.

Reliability of Automated Systems

It has been a general wisdom in the area of human-automation interaction that automation provides a benefit over no automation at about 70% reliability (Lee & See, 2004). Recently laboratory studies have confirmed this assumption. When an automated system has reliability levels lower than about 70%, it is not trusted and the operator treats it as a manual task (Wickens & Dixon, 2007). Recent research (Schuster, 2013) suggested that a priori knowledge of the level of reliability of an automated system may aid performance and reduce inappropriate trust in the system. For the purposes of the current research, the primary visual task will be set at an 80% reliability level, and participants will be informed of the reliability of the system.

Human Attention and Information Processing

As stated previously, much of the research regarding task interruptions has focused on the resumption lag, or the time it takes to resume the primary task after the secondary task ends (Trafton & Monk, 2007). But what if problems associated with interruptions, and even the resumption lag toward the end of the process, are impacted by cognitive processes earlier in the chain of events of task interruptions? Very little research has examined impacts at this end of the spectrum. In order to set the stage for the goals of this research, this section briefly reviews theories of attention, perception, and memory that may be involved early in task interruptions.

Attention and Sensory Stores

The current research relies on assumptions of multiple resource theories (Kahneman, 1973; Wickens, 2002), which have been identified as theories of attention and information processing when multiple sensory modalities and multi-task performance are involved. The current research also relies on theories and laboratory studies regarding the early stages of information processing.

Information processing begins with the sensory registration of a stimulus. Over the years, both psychological and neurological studies have been conducted regarding auditory and visual sensory inputs and their various characteristics. The sensory memory for auditory stimuli is termed “echoic memory”. Studies indicate that the echoic memory store lasts for about 2-5 seconds (Treisman, 1964; Lu, Williamson & Kaufman, 1992; see Cowan, 2000, for review). The sensory memory for the visual stimuli is termed “iconic memory” (Neisser, 1967). Studies of this sensory process have found that iconic memory has duration from about 0.5 to about 1 second (Sperling, 1960).

The research areas regarding the memory stores for both auditory and visual memory as well as auditory working memory are particularly important for the current research, especially in light of theories regarding multiple resources and auditory preemptions.

Multiple Resource Theory

Wickens (1980, 2002, 2008) proposed the multiple resource theory (MRT) to describe and define the attention sharing that occurs when humans process information while engaged in multitasking activities. As described by Wickens (2008), the dangers that can be imposed in operational situations, such as driving or flying, "call for understanding the extent to which such dual-task performance will lead to decreases in time-sharing ability" (p. 449).

Wickens (2008) attributes the conception of his theory to two sources. The first source was the introduction of models of attention and human performance (e.g., Kahneman, 1973; Moray, 1967; and Kalsbeek & Sykes, 1967) that stood in contrast to the selective attention theories by proposing a "general pool of mental 'effort' or undifferentiated resources" for dual-task activities (p. 449). The second source of inspiration was the growing body of literature in the 1950's through early 1970's regarding the effects of divided attention on performance that led to the study of those phenomena as a discipline separate from studies on single modality phenomena. Wickens attributes the inception of this discipline particularly to the works of several authors (Bahrick, Noble, & Fitts, 1954; Bahrick & Shelly, 1958; Briggs, Peters, & Fisher, 1972) and to his own early work in the area (Wickens, 1976).

MRT proposes that in dual or multiple task situations, relative success of performance of the tasks is dependent upon separate information processing resources. That is, task interference can be expected to be greater when the tasks share perceptual modalities (auditory vs. visual channels), visual channels (focal vs. ambient), stages of processing (perceptual/cognitive vs. selection/execution), or processing codes (spatial vs. verbal/linguistic). These modalities, channels, stages, and codes are described next based on descriptions in Wickens (2008).

According to MRT, information acquisition begins with sensing of a stimulus through one of two perceptual modalities—the visual or auditory channel. During performance of dual tasks, when the two tasks are presented separately via these two channels it is referred to as cross-modality time-sharing and the two stimuli do not interfere with each other. Alternatively when two tasks are presented in the same sensory modality, they are less likely to time share and one or both stimuli are either partially or totally masked by the other. One solution is to off-load one task to the other channel which has more resources available.

Recently, the model has added descriptions of two dimensions of the visual channel which are focal (typically central or foveal vision) visual stimuli versus ambient stimuli. Focal stimuli are described as involving object recognition which requires high acuity. Alternatively, ambient vision is distributed across the entire visual field but is most involved in the peripheral vision. Ambient vision involves perception of stimuli such as orientation and movement.

The two stages of processing refer to demands placed on cognition during information processing. For example, a perceptual/cognitive task is the first stage and includes perceiving the stimulus and processing it in working memory. The selection/execution stage would involve selecting an action and executing it based on the previously processed information that passed through perception and working memory.

Processing codes refers to “the distinction between analogue/spatial processes and categorical/symbolic (usually linguistic or verbal) processes” (Wickens, 2002, p. 166). These two types of codes also pass through the same stages of processing—perception/working memory, or selection/execution—as described earlier. Wickens (2002) also notes that these processing codes are often associated with the two hemispheres of the brain and associated with spatial and verbal working memory operations.

To summarize, MRT proposes that humans are able to time share and complete tasks with little cost to overall task performance when those tasks involve different sensory process. For example, auditory interruptions should impose little, if any, costs to the performance of visual tasks because two different pools of information processing resources are being tapped. Several studies have supported this theory (see Wickens, 2002, for a review).

Auditory Preemption

In contrast to the cross-modality time-sharing assumption of MRT, studies have found that discrete auditory interruptions have an alerting characteristic that captures attention and tends to interrupt ongoing tasks whether those tasks are visual, manual, or combined in nature even when the ongoing tasks are of higher priority to overall goals of the operator (Wickens, Sandry, & Vidulich, 1983; Wickens & Liu, 1988; Spence & Driver, 1997; Latorella, 1996, 1998; Strayer & Johnston, 2001; Dismukes, 2001). This phenomenon has called *auditory preemption*.

As proposed, auditory preemption consists of two parts: *onset preemption,* or the immediate alerting and diverting of attention to the auditory interruption, and *strategic preemption*, or the judgments and decisions the operator makes after the initial alerting to attempt to compensate a more complex auditory interruption by determining how to interleave it with a visual ongoing task. It is assumed that strategic preemption is used because, “to do otherwise would risk loss of information from working memory” (p. 464, Wickens, Dixon, &

Seppelt, 2005). There has been little research that has examined the types of strategies used for interleaving auditory and visual tasks.

A number of studies that have attempted to reconcile the cross-modality time-sharing assumption of MRT with auditory preemption theory have met with inconclusive results (Wickens & Liu, 1988; Latorella, 1999; Wickens, Dixon, & Seppelt, 2005). A recent meta-analysis that examined dual task studies that have attempted to evaluate the impact of auditory interrupting tasks on ongoing visual tasks confirmed these inconclusive results (Lu, et al., 2013). Those authors called for continued research to "establish the moderating variables that may tip this balance (between MRT and auditory preemption) one way or the other" (p. 720).

Regarding auditory preemptions, in a study that was designed to examine the concept of Datalink (i.e. ATC communication delivered in an uplinked text-based format), Helleberg and Wickens (2003) used three different display configurations for ATC information: a text-based data link display, a synthesized voice in which pilots also were allowed to write down information on a clipboard thus alleviating overload of working memory, and a display using both types redundantly. A fourth baseline condition was a test of auditory working memory and used the auditory-only, synthesized voice but pilots were not allowed to write down ATC information but required to readback information from memory. For the two auditory-only conditions, the message lengths were varied (2, 3, 4, 5, or 6 discrete instructions per each ATC communication) to study the effects of greater demands on working memory. Four dependent variables were examined: flight path tracking (i.e. root mean square error from the required heading, altitude, and airspeed), visual detection and call-out of conflict aircraft presented after the ATC information (i.e. at 4 seconds, 12 to 24 seconds, or "well after" an auditory alerting tone was given for the ATC information), communication readback errors, and visual scanning data collected via an eye-head tracker system.

Results of the Helleberg and Wickens (2003) study indicated a large effect for flight tracking deviations, showing both a significant cost with longer ATC instructions and a significant main effect of displays with auditory conditions showing the largest flight tracking errors. There was a large main effect for latency in detecting traffic when the ATC information was presented in an auditory format versus the visual format. There was a large main effect for traffic onset such that traffic detection latency decreased when the traffic appeared later in time from the presentation of the auditory ATC instructions. There was a large main effect for readback errors in the ATC auditory display condition versus the visual display condition. Finally, the eye tracking data also indicated more favorable dwell times for the visual display.

The Helleberg and Wickens study offers several useful observations for the current research. For example, effect sizes were large for the impact of auditory communication interruptions on other flight tasks (i.e. flight path tracking—aviating, visual detection of other aircraft—navigating, and readback of ATC instructions—communicating). Also, longer auditory instructions interrupted the flight task more, and errors for detecting critical events outside the aircraft were higher when auditory instructions were presented closer to the "navigate" events.

In relation to the current study, the critical event detections in the Helleberg and Wickens study were for navigation events outside the cockpit (other aircraft) which were discrete events, not ongoing. Also the visual tasks and responses in Helleberg and Wickens differ from the current study. In Helleberg and Wickens, the pilots were required to callout aircraft (detection with verbal response). In the current study, participants monitor an ongoing, automated task inside the interface and choose an appropriate response when errors occur (monitoring and decision-making with motor response). And the current study is designed to present communications that require a response as a secondary task of importance.

Also Helleberg and Wickens (2003) defined complexity in terms of message length. However, preemption theory differentiates onset preemptions (i.e. the initial alerting and diverting of an auditory message) from strategic preemptions (i.e. judgments or decisions an operator must make regarding an auditory message). Thus onset preemptions differ from strategic preemptions in terms of the processing required in working memory (i.e. a judgment or decision). In addition, preemptions may or may not have to do with the length of a message. For example, depending on the information processing requirement of an auditory preemption, a message that has fewer speech elements may be just as taxing as a message containing more elements, depending on the nature and processing requirements of the message. Therefore, further research is needed that evaluates the information processing requirements in working memory for onset and strategic preemptions, not just the span of elements in working memory.

In summary, the current study is designed specifically to test aspects of two theoretical constructs. The first is the impact on the theorized "interruption lag" for secondary communication tasks when those tasks are presented *before* the critical events in the primary, ongoing visual task. The second is a comparison of onset preemptions versus strategic preemptions. For both of these theoretical assumptions, timing of the preempting communication task also is manipulated.

Mental Workload

While there is no one, universally accepted definition of mental workload, the first formal review and attempt to define workload and its measurement is often credited to Moray (1979). Since then, the topic of workload has been extensively studied. Perhaps the closest simple description of mental workload is the concept of "mental effort" or "mental strain", and reflects the "interaction of mental demands imposed on operators by tasks they attend to" (Cain, 2004). Jex (1988) offered the following definition: "Mental workload is the operator's evaluation of the attentional load margin (between their motivated capacity and the current task demands) while achieving adequate task performance in a mission-relevant context" (p. 11).

Workload is considered an inferred construct; that is, it cannot be directly observed and measures of workload and the interpretation of their meanings are inferred from the manipulation of the task difficulty (Wickens, 2001). It is generally agreed that the construct of workload is multidimensional and multifaceted and involves a broad range of situations, time scales, influences, and applications (Jex, 1988).

Cain (2004) stated, "The primary reason for measuring workload is to quantify the mental cost of performing tasks in order to predict operator and system performance" (p. 4-3). Workload measures are often used in human factors and ergonomics laboratory and applied research to compare tasks, or levels of tasks, within a system to each other in order to judge the relative usability among tasks or systems. Four methods of workload measurement have been devised. They are measures of primary task performance, such as under differing task demands; secondary task measures with shedding or degradation in these tasks indicating the primary task is of higher workload; self-report measures, such as direct or indirect operator estimates of their own workload; and physiological or psychophysiological measures (Meshkati, Hancock, Rahimi, & Dawes, 1988).

The most detailed measurement of workload involves convergence and analysis of data from performance, subjective, and physiological measures. However, there are often trade-offs that must be made. The trade-offs take into consideration the relatively lower reliability of subjective reports versus the more sensitive, but much more invasive physiological measurement methods. It has been noted that in practice, subjective measures are used most often, and "these reports seem to be nearly as sensitive and reliable as anything else, and they tend to be far easier to implement" (Flach & Kuperman, 2001, p. 434).

For the current research, the decision was to use a subjective measure of workload based on the NASA-Task Load Index (Hart & Staveland, 1988).

Summary and Research Goals

This review has highlighted several gaps in the literature on communications, task interruptions and auditory preemptions as well as the limitations involved with human information processing. First, conversation interruptions have been found to significantly impact other tasks in multitask systems such as visual tracking, visual detection, vehicle following distance, response time, manual tracking, and response to the conversation (e.g. Strayer, Drews & Johnston, 2003; Angell, et al., 2006; Lu, Wickens, et al., 2013). However, further research is needed regarding the micro-level moments of disruptions of communications to ongoing tasks.

Second, research and theory in the broad area of task interruptions may inform research in aviation communications. Generally, task interruption research has been concerned with the "far end" of a theorized time line for interruptions; that is, on the lag in resumption of the primary task after a secondary interruption (Trafton & Monk, 2007). Far less research has focused on the early stages of task interruption, particularly on the interruption lag which occurs when the operator is preparing to switch from the primary task to the secondary interrupting task. This aspect of the time line may be particularly important when secondary tasks are auditory.

Third, much of the research framed in terms of task interruptions theory has focused on ongoing visual tasks, particularly computer-related tasks, when they are interrupted by another visual task. The ongoing tasks generally do not contain discrete, critical events and can be deferred while the secondary task is handled. Little research has focused on auditory interruptions or conversation tasks and how they impact an ongoing visual task that may contain critical events, such as monitoring an automated system that may not be 100% reliable.

Fourth, multiple resource theory (MRT; Wickens, 2008) proposes that two cross-modal tasks (i.e. one visual, one auditory) time-share cognitive resources and thus may be performed simultaneously with little interference. Some research has supported this cross-modality time-sharing assumption, but other research has not, leading to *auditory preemption* theory (Wickens & Liu, 1988). This theory indirectly proposes a variable of complexity that is different from prior communication research that has studied message length. In addition, a recent meta-analysis has called for continued research to “establish the moderating variables that may tip this balance (between MRT and auditory preemption) one way or the other” (Lu, et al., 2013, p. 720).

Fifth, one aspect of human attention and information processing that has not been directly manipulated in research on auditory preemption theory is sensory memory, particularly echoic memory. Research regarding the early stages of human attention and information processing has found that auditory information has a life of about 2 to 5 seconds in the “echoic” memory store after which it must pass to working memory for rehearsal to retain it, or decay. This echoic memory limitation to human information processing may be at the heart of the question regarding auditory preemptions and deserves further explanation.

Onset preemptions divert attention only for a short time, possibly because the only information they carry is an alert or reminder to the operator to perform some task that is already well-rehearsed (i.e. stored in long-term memory). In that way, an onset preemption may be nothing more than an auditory cue to perform a well-learned process (e.g. push a button, read a gauge, repeat a word) that does not require new information processing and so does not compete with performance such as detections of critical events in the ongoing visual task.

On the other hand, strategic preemptions have not only an alerting function (because the human system is wired to perceive auditory stimuli in that way), but also require further information processing in order to be resolved. It is possible that information in the auditory channel for strategic preemptions cannot simply be attached to some other external information or process in long-term memory in the first 2 to 5 seconds. This new information must be dealt with in working memory which would require a strategy for its completion. In this way, strategic preemptions could be expected to compete with performance of the ongoing visual task, and critical events in the task may be missed. Thus, it seems important to consider timing of interrupting communications—both simple and complex—in order to further evaluate complexity *and* timing as moderators that may differentiate MRT from auditory preemptions.

Considering these research gaps—i.e. what strategies are used to interleave simple (onset) versus complex (strategic) auditory preemptions with visual tasks; how the timing of communication interruptions impact visual tasks; and whether such complexity and timing moderators during the task switching process explain when auditory and visual tasks time-share cognitive resources and when they compete—a revised time line for the study of communication interruptions is presented in Figure 2 that will guide the current research.

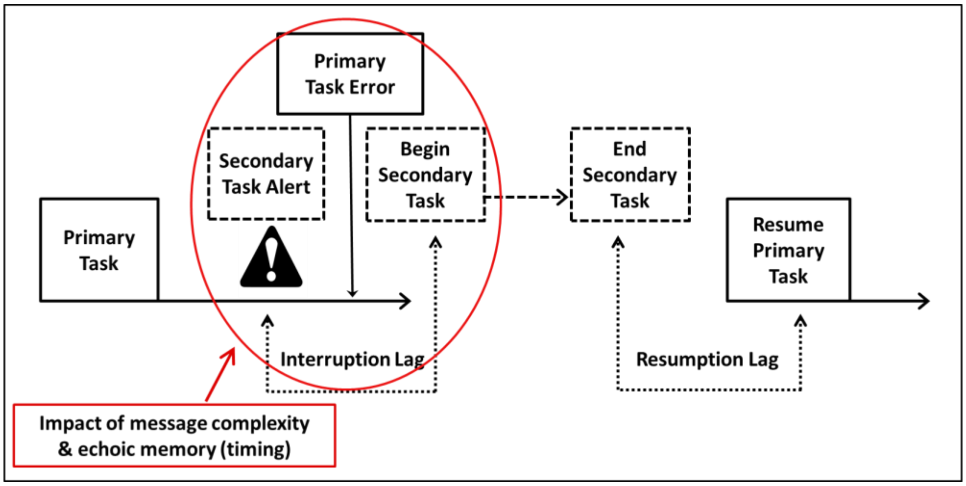

Figure 2. Revised task interruption time line highlighting impacts of message complexity and timing at the task-switching interruption lag

There are two primary goals for this research. The goal of Study 1 is to evaluate the impact of complexity and timing of communications when they occur before or simultaneous to a critical event in an ongoing visual detection task (i.e. the system monitoring task). The goal of Study 2 is to evaluate the impact of complexity and timing of communications when they occur after a malfunction in an ongoing visual detection task (i.e. the system monitoring task). This study was conducted to test alternative assumptions that some other factor, perhaps anticipation of communication events, would impact performance in the visual system monitoring task.

In this research, both complexity and timing were examined in the two studies, with the placement of the communications manipulated between the studies (i.e. before or after a visual detection task). Prior studies that have defined communication complexity in terms of message length have likely mixed the onset versus strategic communication types within a single message. For example, one instruction in a message containing five total instructions might include an immediate alert that a message is occurring (e.g. “Citation One Four Six Charlie”), an element that requires spatial memory (e.g. “turn left heading three six zero”), an informational element such as hazard alert (e.g. “traffic in the pattern is a Cessna”), an element that requires a manual response (e.g. “contact approach one twenty two point seven”), or an implied requirement to acknowledge that the message was received (e.g. repeating one’s call sign at the end of the message readback), among other types. These prior studies have not examined complexity among instruction types. The present research defines complexity in terms of onset (i.e. a simple, repetitive alerting message) versus strategic (i.e. a message requiring working memory processes) preemption message types.

To examine the timing variable, the communication events in this research were varied based on the time, in seconds, that a communication request occurred relative to a system monitoring malfunction, and based on research regarding task interruptions and sensory memory stores. Timings before critical malfunction events (Study 1) were chosen in order to evaluate the differential impact of a simple communication task (which should require no more than about 5 seconds for a response) versus a complex communication task (which will require longer than 5 seconds due to activation of working memory process, but likely less than 15 seconds) where processing time overlaps the presentation of the critical visual event. The communications were purposely kept short based on prior task interruption research that has indicated that interruptions of longer than 13 seconds create an exponential impact on the length of the resumption lag.

In Study 2, the placement of a visual system monitoring event was varied at time increments, in seconds, like those in Study 1. Thus if the temporal placement of communication events in relation to visual detection tasks is not a factor in the detection of visual malfunctions, then the results of Study 1 and Study 2 should be similar. Figure 3 illustrates the overall conceptualization of this research.

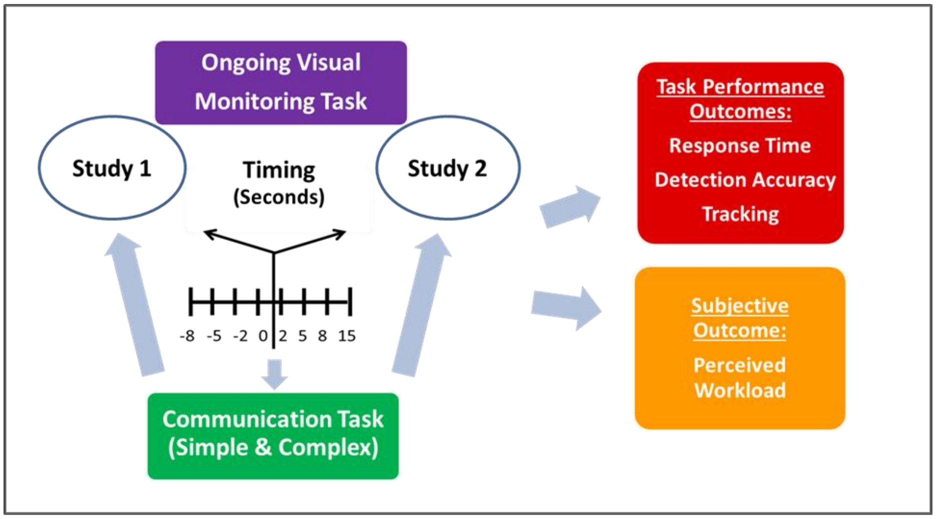

Figure 3. Current research conceptualization.

CHAPTER THREE: METHODOLOGY

Research Design

This research consisted of two studies. A 2 (communication complexity) x 4 (communication timing interval) within subjects design was used for each study. The studies were defined as communications occurring before critical events in a system monitoring task (Study 1) and after those events (Study 2). The overall research design is represented in Table 1. As indicated in the research design table, each study consisted of eight experimental blocks, four blocks for each level of communication complexity by timing of the presentation of communications.

Table 1

Research design with study and independent variable manipulations

Placement of Communication Events	Communication Complexity Manipulation	Communication Timing Manipulation
Study 1 Before system malfunctions	Simple	8 seconds before 5 seconds before 2 seconds before Simultaneously
	Complex	8 seconds before 5 seconds before 2 seconds before Simultaneously
Study 2 After system malfunctions	Simple	2 seconds after 5 seconds after 8 seconds after 15 seconds after
	Complex	2 seconds after 5 seconds after 8 seconds after 15 seconds after

The conditions (blocks) in each of the studies were defined by both complexity of the communication and the timing, in seconds, when the communication occurred based on limitations of sensory memory stores. In Study 1, the timing intervals were simultaneous to, or 2, 5, or 8 seconds before the visual detection event. Therefore, the eight timing blocks were referred to as -8 Simple, -5 Simple, -2 Simple, 0 Simple, -8 Complex, -5 Complex, -2 Complex, and 0 Complex. The purpose of Study 1 was to evaluate the limits of echoic memory under auditory preemption theory and considering the concept of the interruption lag in theorized task interruption timelines.

In Study 2, the timing of communications events occurred at 2, 5, or 8 seconds after the visual detection event, and a fourth condition was placed well outside these time limits, at 15 to 20 seconds after a visual detection event. This final "15+" condition is intended to act as a baseline condition in which the auditory and visual tasks would not be expected to compete. Therefore, the eight timing blocks were referred to as 15 Simple, 8 Simple, 5 Simple, 2 Simple, 15 Complex, 8 Complex, 5 Complex, and 2 Complex. The purpose of Study 2 was to test the alternative assumption that some other factor, perhaps anticipation of communication events, would impact performance of study tasks. Study 2 results were also expected to provide information regarding the limits of iconic (visual) memory as well as consideration of the interruption lag in task interruption timelines, similar to Study 1.

The blocks were 7-minutes long for a total of 56 minutes in study tasks. After each 7-minute timing block, the system displayed the workload rating scale for the participant to complete regarding the workload for the block they had just experienced. Further details of the set-up for each study task can be found in the apparatus section. Considering this research design, the following goals and hypotheses were generated for the two studies.

Study 1 Research Goal

The goal of Study 1 was to evaluate the impact of complexity and timing of communications when they occur before or simultaneous to a malfunction in an ongoing visual detection task (i.e. the system monitoring task). Considering auditory preemption theory and the limitations of the auditory sensory store, it was expected that a communication that occurs simultaneous to or within 5 seconds before a visual detection task would be most disruptive to the operator, with complex communications leading to worse performance outcomes and higher subjective workload compared to simple communications. Beyond 5 seconds, these impacts were not anticipated. Given these theoretical assumptions, the following hypotheses were proposed.

Study 1 Hypotheses

Study 1 task performance hypotheses.

Hypothesis 1. Response time to system malfunctions will be longer for complex communications compared to simple communications (a simple effect).

Hypothesis 2. The onset of communications will delay response times to system malfunctions depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 2a. Response times to system malfunctions will be longest when communications occur simultaneously with or closest to system malfunctions with the longest response times occurring for complex communications (a communication complexity by timing interaction effect).

Hypothesis 2b. Response times to system malfunctions will not be impacted among conditions where either simple or complex communications occur at 8 seconds before the malfunctions (a null effect that tests the limits of echoic memory).

Hypothesis 3. Percent of correct detections of system malfunctions will be lower for complex communications compared to simple communications (a simple effect).

Hypothesis 4. The onset of communications will lead to poorer detections of system malfunctions depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 4a. Detection of system malfunctions will be lowest when communications occur simultaneously with or closest to system malfunctions with the worst rates of detection occurring for complex communications (a communication complexity by timing interaction effect).

Hypothesis 4b. Detection of system malfunctions will not be impacted in conditions where either simple or complex communications occur at 8 seconds before the malfunctions (a null effect that tests the limits of echoic memory).

Hypothesis 5. A concurrent, manual tracking task will be performed with more error in the presence of complex communications compared to simple communications (a simple effect).

Hypothesis 6. The onset of communications will lead to more tracking task error depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 6a. Tracking error will be highest in conditions where communications occur simultaneously with or closest to system malfunctions with the highest tracking error occurring for complex communications (a communication complexity by timing interaction effect).

Hypothesis 6b. No differences will be observed for tracking task accuracy among the conditions where either simple or complex communications occur at 8 seconds before system malfunctions (a null effect that tests the limits of echoic memory).

Study 1 subjective workload hypotheses.

Hypothesis 7. Workload will be perceived as higher for complex communications compared to simple communications (a simple effect).

Hypothesis 8. The onset of communications will result in higher subjective workload depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 8a. Subjective workload will be highest when communications occur simultaneously with system malfunctions, followed by 2, 5 and 8 seconds before the malfunctions, with the highest workload occurring in the presence of complex communications at these timing intervals (a communication complexity by timing interaction effect).

Hypothesis 8b. No differences will be observed for subjective workload among the conditions where either simple or complex communications occur at 8 seconds before system malfunctions (a null effect that tests the limits of echoic memory).

Study 2 Research Goal

The goal of Study 2 was to evaluate the impact of complexity and timing of communications when they occur after a malfunction in an ongoing visual detection task (i.e. the system monitoring task). The timing and complexity of communications were similarly manipulated as in Study 1. This study was conducted to test an alternative assumption that any significant findings in Study 1 were not due to the impact of communications on visual detection tasks, but perhaps some other factor such as distraction by anticipation of an upcoming communication event. If some other factor was in operation, then the results of Study 1 and Study 2 were expected to be similar.

Based on the limits of iconic memory, a visual stimulus has a life of only about 1 second. Therefore, if some other factor was not responsible for the findings in Study 1, then it was expected that participants would respond to the system malfunctions first before responding to the communication requests. Thus no interaction effects were anticipated. However, based on research that communications, in general, disrupt various types of transportation-related tasks, main effects were anticipated for simple versus complex communications on study dependent variables.

Study 2 Hypotheses

Study 2 task performance hypotheses.

Hypothesis 1. Response times to system malfunctions will be longer in conditions with complex communications compared to simple communications conditions, and no interaction of complexity and timing is anticipated (a simple effect).

Hypothesis 2. Percent of correct detections of system malfunctions will be lower for complex communications compared to simple communications conditions, and no interaction of complexity and timing of communications is anticipated (a simple effect).

Hypothesis 3. A concurrent, manual tracking task will be performed with more error in the presence of complex communications compared to simple communications, and no interaction of complexity and timing of communications is anticipated (a simple effect).

Study 2 subjective workload hypothesis.

Hypothesis 4. Workload will be perceived as higher for complex compared to simple communication conditions, and no interaction of complexity and timing of communications is anticipated (a simple effect).

Participants

Participants were recruited through the University of Central Florida (UCF) psychology research recruitment system and compensated with course credit for their participation. The protocol for this research was approved by the UCF Institutional Review Board (IRB) prior to data collection. A copy of the approval is contained in Appendix A. Prior studies of response times for conversation interruptions indicated a medium to large effect size could be expected. An a priori power analysis was conducted using the G*Power 3 computer program (Faul et al., 2007). The parameters included an estimated medium effect size (0.25), alpha of .05, a desired power of 0.80, two groups, four repetitions, and a correlation between repeated measures of .50. The power analysis estimated total sample size of 24 per experiment, 48 in total.

A total of 52 persons were recruited for the two studies over a six-week period of time. Participants were randomly assigned to studies and conditions within studies. All participants were given information regarding the research and consented to participate. All participants who started the study eventually finished. However, data from four participants had to be excluded, two due to software malfunctions, and two due to obvious random keyboard entries on the main study task of interest (the system monitoring task).

The final sample consisted of a total of 48 participants that included 20 males and 28 females ranging in age from 18 to 36 (M = 21.06, SD = 3.83). All reported 20/20 vision, corrected or uncorrected. None reported color blindness. Five reported left-handedness and all spoke English proficiently. Daily computer use was reported by 43 of the 48, with the others reported computer use several times a week. Forty-three participants rated themselves as intermediate in computer skills, and the remaining five reported expert status.

Apparatus and Materials

The Multiple Attribute Task Battery II (MATB-II, NASA-LARC, 2011) was used as the research platform. Four tasks of the MATB-II were used as described below. The experiment apparatus is shown in Figure 4.

Two identical experiment stations were used. Participants were run two at a time when the study slots were filled. Each experiment station consisted of the MATB-II which was run on a standard laptop computer with a color monitor display attached peripherally. Participants used a standard mouse, keyboard and two-axis joystick to make inputs for the study tasks. Each of those peripheral devices was connected to the laptop computer through USB ports. They listened to communications over a headset that offered a moderate level of noise attenuation. The headphones received output from the laptop via the computer's audio jack. All peripherals were identical in make and model for the two experiment stations. A MATB-II interface is shown in Figure 5.

Based on prior research regarding instructions about task priorities and system reliability (Wickens & Dixon, 2007), all participants were told to give equal priority to all visual tasks (monitoring, tracking, resource management) and to listen and respond to the communications. How they decided to attend to and accomplish these tasks was at their discretion. They were also told about the system reliability.

Each participant completed eight, 7-minute long communication timing interval blocks. Each block represented one of the two levels of communication complexity (simple or complex) and one of the four timing intervals (communications presented at 0, -2, -4 or -8 seconds before a system monitoring malfunction in Study 1, or at 2, 5, 8 or 15 seconds after the malfunction in Study 2).

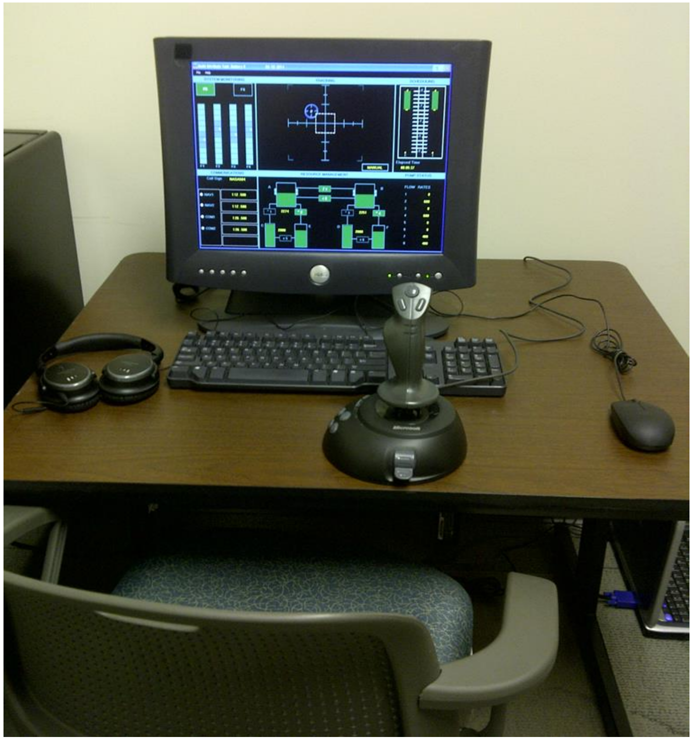

Figure 4. Experiment apparatus.

Monitor, keyboard, mouse, two-axis joystick, and headphones; the MATB-II was run on laptop computers that are not shown.

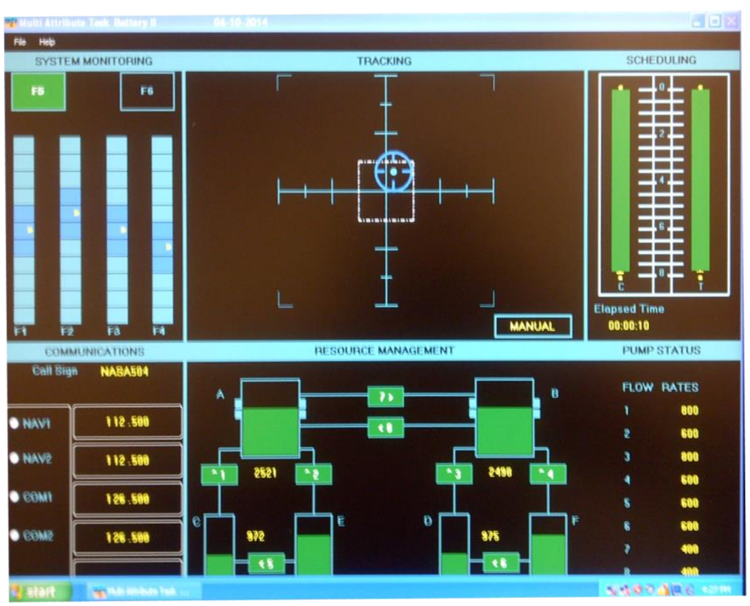

Figure 5. The Multi-Attribute Task Battery-II interface.

System Monitoring Task and Manipulation

The system monitoring task (SYSMON) was the primary, visual task of interest in this study. It consists of four gauges each with a pointer that normally fluctuates around the center of the gauge. A "malfunction" was indicated when the pointer offset to the extreme top or bottom of the gauge. The offsets were coded to last for 10 seconds after which the pointer returned to fluctuate around the center. Participants were not told about this duration.

Participants were instructed that the system was automated but as the operator their task was to monitor the four system gauges and to "fix" any critical malfunctions when they occurred by pressing designated keys (F1 to F4) to return the pointer to its normal range. The gauge offsets varied across each block in order to appear randomly spaced.

As suggested by Wickens and Dixon (2007), about 70% is the minimum level of reliability for diagnostic aiding automation where benefit is achieved beyond manually performing the task. In addition, telling operators about system reliability assists in forming appropriate trust in the system. Therefore, the MATB-II event files were scripted to present an 80% reliable system and participants were told about the reliability level of the system.

The SYSMON task was designed to deliver 10 gauge offsets in each 7-minute timing block of the research design. Of those gauge offsets, seven were placed to coincide with a communication event for that timing block. An additional three gauge offsets and three communication events occurred outside of the timing interval for the block so that participants would not associate the SYSMON malfunctions only with communication events.

The SYSMON log files recorded both the onset and the timing of the SYSMON events as well as the participant key presses and response time to the events. The data from the logs became the accuracy and timing measures for this study. To attempt to prevent participants from randomly pressing buttons to try to increase the chances of catching these errors, they were told that they would also need to monitor and respond to the communications over the headsets, attend to the resource management task, and maintain the tracking symbol in the middle of that window. They had been told in advance that this was a multi-tasking study. During the training, no one task was emphasized over any other.

In terms of Rasmussen's (1983) levels of behavior in human performance, this task is conceptualized as a knowledge-based task. Participants must develop and maintain a mental model of the functioning of the system, and they must make decisions across time regarding interventions to system functioning.

Tracking Task

The tracking task (TRACK) represents a manual flight navigation task in which the operator watches a specified area on the screen and maintains a symbol in the center of that area. This task was performed manually during the entire experiment with a standard USB joystick. This was the secondary performance task of interest and also functioned as a primary distractor task in the physical modality. The root mean square error (RMSE) of the target from the center of the task window was recorded in one second intervals by the system and written to the participant's TRACK log file. That data from the log files became the tracking task measure for this study. Tracking is considered a skill-based behavior in Rasmussen's (1983) model of human performance.

Resource Management Task

The resource management task (RESMAN) represents an in-flight fuel management task. Participants were given the instruction that they should maintain two "fuel tanks" within an optimum range by transferring fuel to each of the tanks from two corresponding primary and secondary supply tanks through designated "pumps". This task functions as a primary distractor task in the visual modality, and a rule-based behavior in Rasmussen's (1983) model.

Communication Task and Manipulation

The communications task (COMMS) was the primary distractor in the auditory modality and can be conceptualized as a rule-based task under Rasmussen's (1983) model of human performance. The dependent variables of interest in this study examined the effects of the communication task on operator performance (i.e. SYSMON accuracy and response time, and the TRACK RMSE) and workload.

In the simple communication events, participants were required to repeat three, pre-recorded words that they heard over the headsets. For example, the participant may have heard, "Repeat: Skillet, chiefs, owner" and they were to repeat those words exactly.

In the complex communication events, participants were required to think of three words that began with the same letter as the last letter of the word that they heard as a prompt. For example, the participant might hear, "Say three words starting with the last letter in shrimp" to which a participant might say, "Pig, pole, pond."

The choice of either repeating or generating three words as representations of the two levels of communication complexity (i.e. simple or complex) was adapted from procedures in Strayer and Johnston (2001) and represent different levels of information processing complexity. The use of an information processing task to represent communications in experimental research was supported in a recent meta-analysis (Horrey & Wickens, 2006) and have been used in larger studies to examine the impact of communications on transportation tasks (Angell, et al., 2006).

The communication events were recorded and then programmed into the MATB-II system scripts at the specified timings for each block. All pre-recorded words were spoken by a female. Participants were not aware of the number or frequency of presentation of the words across the study.

The list of words for this task, both those that required repeating and those that were given as prompts, were taken from 6^{th} grade spelling word lists on educational websites. Only words of between four and seven letters were chosen from the word lists. A 6^{th} grade level was used in order to provide adequate word difficulty and equivalence across blocks. The participant's verbal responses were collected offline by the experimenter with the use of the communication task observation sheet. The words can be found in the communication task observation sheet in Appendix E.

Measures

Demographics

A demographic questionnaire (see Appendix D) was administered after the training session for the study. It included questions regarding age, gender, hours spent using the computer per week, self-ratings of level of expertise with computers (i.e. novice, intermediate, expert), frequency and hours per week playing video games, and frequency of playing aviation-related games.

Performance Measures

Performance measures were collected the MATB-II system. For the system monitoring task, measures included correct detection of the system monitoring malfunctions, expressed as a percentage, and response times to the errors. An event log for the SYSMON data can be found in Appendix F.

Tracking task performance was measured as root mean square error (RMSE) from the center of the target grid. Resource Management task data is collected by the system, but was not further analyzed for this study.

Communication Task Measures

The original research plan included counting the frequencies of correct responses to the communication events for later analysis. However, the data collected via the check sheets indicated that nearly all participants responded to the communications correctly. There was a near ceiling effect for this data, thus it offered little useful information for further analysis.

It was not possible with the current apparatus to record participant verbalizations and then matches those to the timing of performance of tasks in the MATB-II with any degree of precision. Therefore, response time data for the communications task was not captured.

Workload

The NASA Task Load Index (NASA-TLX) was administered by the MATB-II system following each 7-minute block for a total of eight workload ratings for each participant. The data was output to an Excel spreadsheet by the system.

Procedure

Participants performed all eight communication complexity by timing conditions (blocks) for one of the two studies. They were randomly assigned to one study and one variation of the sequence of presentation of the complexity by timing conditions based on a row of an order 8 Latin Square design. The Latin Square design sequences for Study 1 and 2 are presented in the figures below.

A	-8-S	-5-S	0-C	-2-S	-2-C	0-S	-5-C	-8-C
B	-5-S	-2-S	-8-S	0-S	0-C	-8-C	-2-C	-5-C
C	-2-S	0-S	-5-S	-8-C	-8-S	-5-C	0-C	-2-C
D	0-S	-8-C	-2-S	-5-C	-5-S	-2-C	-8-S	0-C
E	-8-C	-5-C	0-S	-2-C	-2-S	0-C	-5-S	-8-S
F	-5-C	-2-C	-8-C	0-C	0-S	-8-S	-2-S	-5-S
G	-2-C	0-C	-5-C	-8-S	-8-C	-5-S	0-S	-2-S
H	0-C	-8-S	-2-C	-5-S	-5-C	-2-S	-8-C	0-S

Block	Timing (Seconds)	Complexity
1 = -8-S	-8	Simple
2 = -5-S	-5	Simple
3 = -2-S	-2	Simple
4 = 0-S	0	Simple
5 = -8-C	-8	Complex
6 = -5-C	-5	Complex
7 = -2-C	-2	Complex
8 = 0-C	0	Complex

Figure 6. Order 8 Latin Square design and legend explaining timing blocks for Study 1. (The “0” blocks are full pairing of communications with system monitoring malfunctions.)

A	8-S	5-S	0-C	2-S	2-C	0-S	5-C	8-C
B	5-S	2-S	8-S	0-S	0-C	8-C	2-C	5-C
C	2-S	0-S	5-S	8-C	8-S	5-C	0-C	2-C
D	0-S	8-C	2-S	5-C	5-S	2-C	8-S	0-C
E	8-C	5-C	0-S	2-C	2-S	0-C	5-S	8-S
F	5-C	2-C	8-C	0-C	0-S	8-S	2-S	5-S
G	2-C	0-C	5-C	8-S	8-C	5-S	0-S	2-S
H	0-C	8-S	2-C	5-S	5-C	2-S	8-C	0-S

Block	Timing (Seconds)	Complexity
1 = 8-S	8	Simple
2 = 5-S	5	Simple
3 = 2-S	2	Simple
4 = 0-S	0-no pairing	Simple
5 = 8-C	8	Complex
6 = 5-C	5	Complex
7 = 2-C	2	Complex
8 = 0-C	0- no pairing	Complex

Figure 7. Order 8 Latin Square design and legend explaining timing blocks for Study 2. The “0” blocks are no pairing of communications with system monitoring malfunctions.

The study was advertised as “Conversations in Multitasking Environments” in order to emphasize the focus on both conversing and on multitasking. Participants were greeted upon arrival and given a copy of the informed consent for review. The consent form is contained in Appendix B. After they read the consent, the study and its tasks and equipment were briefly explained and any questions were answered.

After the experiment explanation, participants were trained in the use of the system using a 7-minute training script in the MATB-II. However, the system was paused after each task in order to explain the next task, thus the entire training lasted about 25 minutes. After training, participants completed the demographics questionnaire. The study immediately followed the training without breaks. After the study, participants were thanked and credits were awarded in the department research system. The full experimenter script for the study can be found in Appendix C. The timeline for both studies is shown in Table 2.

Table 2

Study timeline

Time from start	Study event
0:10	Study explanation and consent process
0:35	Training and practice of experiment tasks; demographics questionnaire
1:45	Experiment: 8 blocks of 7 minutes + 1 minute for workload survey
1:50	Debrief and study exit

CHAPTER FOUR: RESULTS AND DISCUSSION

This chapter presents results from Study 1 and Study 2. The IBM SPSS Statistics, Version 22 statistical analysis package was used to analyze the study results. Unless otherwise specified, all hypotheses were tested at alpha of .05.

Prior to analysis of the results by study, a check of random assignments to the two studies and manipulation checks for the timing and complexity of the presentation of communications were conducted. Those results are presented first.

Demographic Variables for the Two Studies

Means, standard deviations, and intercorrelations among the demographic variables for the combined Study 1 and Study 2 data are presented in Table 3.

Table 3

Descriptive statistics and intercorrelations among study variables for two studies combined

Variable	*M*	SD	1	2	3	4	5	6
1. Age	21.06	3.83	-					
2. Gender	1.58	0.50	.01	-				
3. Hours on comptuer per week	2.79	1.27	.18	.06	-			
4. Confidence with video games	3.25	1.06	-.14	-.44**	.09	-		
5. Weekly hours playing games	1.25	0.60	-.05	-.36*	0.26	.50**	-	
6. Frequency of aviation game play	1.33	0.63	.10	-.16	.09	.25	.39**	-

** Correlation significant at 0.01 level (2-tailed). * Correlation significant at 0.05 level (2-tailed).
Gender: Male = 0, Female = 1. Questions 3 - 6 ratings: 1 (low/0-9 hours/never) to 5 (high/40+ hours/daily).

Four significant correlations were found. Males were significantly more confident in their video game skills and spent more hours playing video games per week than females. Also those who played aviation-related video games played more video games in general than those who did not play aviation-related video games. Hours of game play per week was also correlated with confidence in video game play.

Check of Random Assignment to the Two Studies

A series of one-way analyses of variance (ANOVAs) were conducted on the demographic variables to assess whether random assignment of participants to the two studies was achieved. The means and test statistics are listed in Table 4.

Table 4

Group means and standard deviations for demographic variables by study

Variable	Overall *M (SD)*	Study 1 *M (SD)*	Study 2 *M (SD)*	*df*	*F*	*p*
1. Age	21.06 (3.83)	20.5 (3.07)	21.63 (4.47)	1, 46	1.03	.32
2. Hours on comptuer per week	2.79 (1.27)	2.88 (1.42)	2.71 (1.12)	1, 46	0.21	.66
3. Confidence with video games	3.25 (1.06)	3.25 (1.11)	3.25 (1.03)	1, 46	0.00	1.0
4. Weekly hours playing games	1.25 (0.60)	1.21 (0.59)	1.29 (0.62)	1, 46	0.23	.64
5. Frequency of aviation game play	1.33 (0.63)	1.33 (0.57)	1.33 (0.70)	1, 46	0.00	1.0

Examination of the ANOVA statistics indicated that there were no differences between the groups in the two studies on the demographic variables of age, hours spent on the computer per week, confidence with video games, weekly hours playing video games, and frequency of aviation games play. These analyses suggest that random assignment of participants to conditions was achieved.

Manipulation Checks for Communication Complexity

Checks were performed to assess whether the manipulation of complexity of communications was achieved. A series of t-tests were conducted to compare the grand mean of the simple versus the grand mean of complex communication conditions, regardless of study, on the four primary dependent variables in the study. Results are presented in Table 5.

Table 5

Communication complexity manipulation checks for study dependent variables

Dependent variable	Simple Communications *M (SD)*	Complex Communications *M (SD)*	*t*	*df*	*sig. (2-tailed)*
Detections Accuracy	89.88 (10.03)	85.34 (14.25)	2.59	47	.01
Detections Response Times	4.18 (1.30)	4.62 (1.48)	-2.65	47	.01
Tracking Task Accuracy	29.37 (6.78)	29.47 (7.54)	.22	47	.82
Pereived Workload	42.25 (17.24)	48.23 (16.01)	-5.24	47	.0001

Results indicated that the manipulation of the complexity of communications (i.e. simple versus complex) was achieved. The t-tests comparing the mean values for the simple versus the complex communication blocks on the detection of system monitoring malfunctions, response times to the malfunctions, and subjective workload were all highly significant for the 2-tailed significance tests.

However, the tracking task accuracy variable was not significant when comparing the simple versus the complex conditions. Several factors may contribute to this null finding.

First, it is possible that participants always maintained accuracy for this task regardless of communication or the system monitoring events that occurred. Since this was the only fully manual task, the “hands-on” nature of the task may have commanded participants’ full attention.

There may also be a software explanation. Several participants reported a lag in the inputs of the tracking task as well as a relative ease in performing this task. That is, the task was not highly sensitive to inputs and often participants could take hands off the task for several seconds before inputs were required to maintain the target within the specified box.

Finally, a statistical explanation is possible. The values for the root mean square error from center for the tracking task data was averaged over the entire timing block rather than only near the communication presentations. This averaging method may have introduced noise in the data that may have obscured any differences. Later analyses of the tracking data may help to determine which of these factors may have influenced these results.

Tests of Normality for the Two Studies

Tests of normality for each of the four dependent variables were conducted for each study independently. The value used for each dependent variable was the mean of the scores across the four simple and the four complex communications conditions. The test value was computed as the skew value to standard error of skew. A significant skew value was defined as any ratio greater than an absolute value of 2. Results are presented in Table 6.

Table 6

Tests of normality for dependent variables by communications complexity and study

Measure	Level of Complexity	Study 1 Skew Ratio	Study 2 Skew Ratio
Detections Accuracy (Percents)	Simple	-1.51	-5.39
	Complex	-1.96	-2.21
Detections Response Times	Simple	0.87	1.74
	Complex	0.64	0.25
Tracking Task Accuracy	Simple	1.10	1.12
	Complex	1.06	1.84
Perceived Workload	Simple	-0.58	-0.99
	Complex	-1.49	-2.27

Using the mean of the scores across the four simple and the four complex communication conditions by study, three of the 16 skew ratios were found significant. They were percent of correct detections of system malfunctions for simple and complex conditions in study 2, and workload ratings for the complex communication conditions in study 2. However, two of those three values were within only tenths of a point of the cut-off value of 2. Therefore, given these relatively good values for skew, the decision was made not to transform the data.

Study 1: Communications Occurring Before System Malfunctions

Study 1 Research Goal

The goal of Study 1 was to evaluate the impact of complexity and timing of communications when they occur before or simultaneous to a malfunction in an ongoing visual detection task (i.e. the system monitoring task).

Tests of Hypotheses

Tests of the main hypotheses for Study 1 were conducted using a series of 2 (communication complexity: simple and complex) x 4 (communication timing: 0, -2, -5, and -8 seconds) factorial repeated measures ANOVAs. Several sub-hypotheses were evaluated using paired samples t-tests with 2-tailed tests of significance. Estimates of effect sized are based on those provided in the G*Power program (Faul, Erdfelder, Lang, & Buchner, 2007) as well as in Pallant (2007) and Cohen (1988). Graphics accompany the hypotheses, as appropriate.

Hypothesis 1. Response time to system malfunctions will be longer for complex communications compared to simple communications (a simple effect).

The response time variable was the time the participant took to respond to system monitoring malfunctions when they were paired with both simple and complex communications that required responses. The response time values ranged from 0 to 10 seconds, measured to four decimal places. These times were averaged across each of the eight, 7-minute timing blocks as described above. The participants' values for the eight blocks were then submitted to a 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA.

To test this hypothesis, the response times were evaluated with the main effects test for complexity. This test was significant, F (1, 23) = 7.13, p = .014; partial $\eta 2$ = .24, a medium effect. Participants performed significantly better in the simple (M = 4.50, SD = 1.11) compared to the complex (M = 5.15, SD = 1.60) communication task.

For Study 1, the response times to system malfunctions, on average, were about two-thirds of a second longer in the complex conditions compared to the simple communication conditions.

Hypothesis 2. The onset of communications will delay response times to system malfunctions depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 2a. Response times to system malfunctions will be longest when communications occur simultaneously with or closest to system malfunctions with the longest response times occurring for complex communications (a communication complexity by timing interaction effect).

The hypothesized order of the effect for the longest to the shortest response times was 0, -2, -5, and -8 seconds complex communications, followed by 0, -2, -5, and -8 second simple communication conditions. Response times were expected to follow a linear pattern with the longest times occurring in the time blocks where the communication and system malfunction were the closest. Complex communications were expected to elicit the longest response times.

The means, standard deviations, and standard errors for the response times for each of eight time blocks are presented in Table 7.

Table 7

Means, standard deviations, and standard errors for response times when communications occur before system malfunctions

Timing Interval & Complexity	Mean	SD	SE
0 second complex	5.91	2.07	0.42
-2 second complex	5.35	1.89	0.39
-5 second complex	4.90	1.97	0.40
-8 second complex	4.44	1.64	0.33
0 second simple	4.84	1.58	0.32
-2 second simple	4.20	1.72	0.35
-5 second simple	4.29	1.82	0.37
-8 second simple	4.66	1.27	0.26

The main effect for complexity was reported above. The main effect of timing on response time also was significant, F (3, 69) = 4.84, p = .004; partial $\eta 2$ = .17, a small to medium. Participants' response times were significantly slower in the simultaneous (M= 5.38) than the -2 second (M = 4.78), -5 second (M = 4.59) and -8 second (M = 4.55) timings. The interaction of complexity and timing on response time was significant, F (3, 69) = 2.73, p = .05; partial $\eta 2$ = .11, a small effect.

The post hoc comparisons for the interactions indicated significant differences between the simple (M = 4.84; SD = 1.58) and complex (M = 5.91; SD = 2.07) simultaneous communications, and between the simple (M = 4.19; SD = 1.72) and complex (M = 5.35; SD = 1.89) -2 second conditions. A linear trend for the complex communication conditions can be seen in Figure 8.

Examination of the means table shows that compared to the presentations of simple communications, participants required about 0.5 to 1.7 seconds longer to respond to a system monitoring malfunction when the malfunction was paired with a complex communication or when the complex communication occurred at -2 seconds before the malfunction.

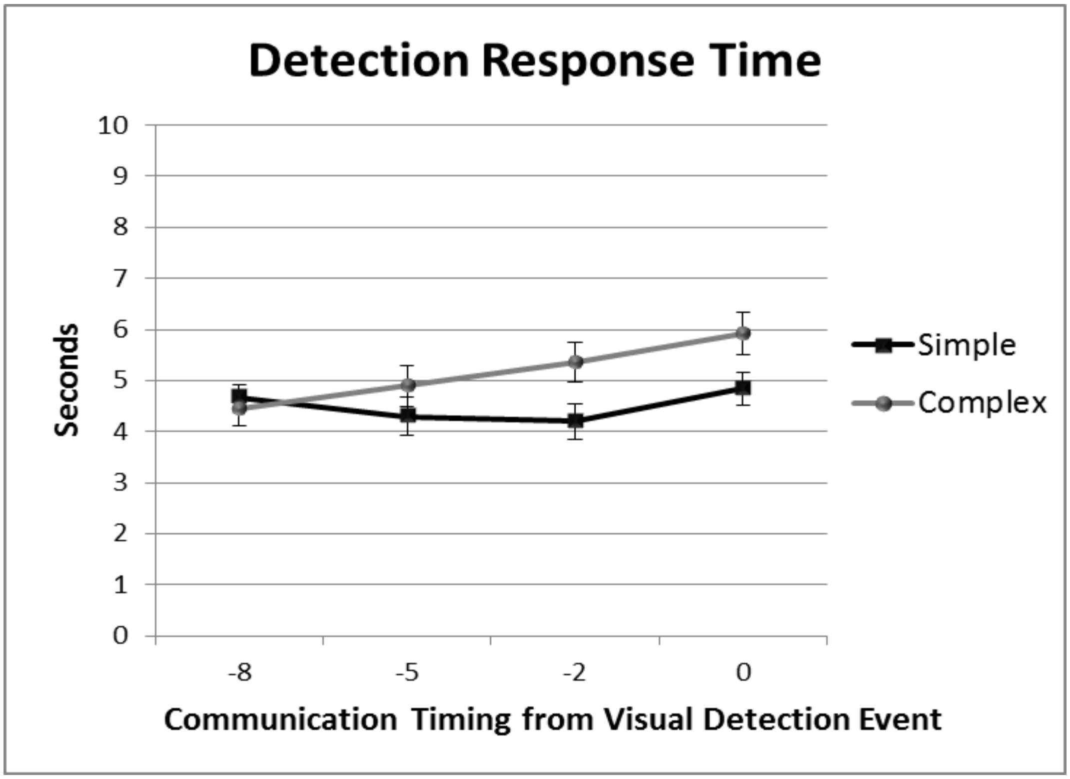

Figure 8. Response times when communications occur before system malfunctions.

Hypothesis 2b. Response times to system malfunctions will not be impacted among conditions where either simple or complex communications occur at 8 seconds before the malfunctions (a null effect that tests the limits of echoic memory).

This secondary hypothesis for response times was assessed with a paired-samples t-test to compare the two, -8-second communication conditions (simple and complex) in Study 1. The results supported the hypothesis. There was not a statistically significant difference when comparing the -8 second simple (*M* = 4.66, SD = 1.27) and the -8 second complex (*M* = 4.44 SD = 1.64) communication conditions on the variable of response time to system malfunctions, t (23) = .51, p = .62 (two-tailed), a very small effect, η^2= .01. The relationship between the two conditions can also be seen in Figure 8.

Hypothesis 3. Percent of correct detections of system malfunctions will be lower for complex communications compared to simple communications (a simple effect).

The percent of correct detections variable was the number of detections and resets of the four system monitoring gauges versus the total number of offsets within a time block. The gauge offsets were paired with either simple or complex communications that required responses in each of four time blocks. The percent of correct detections per block was the average across each of the eight, 7-minute timing blocks (simple and complex). The possible range of values was 0 to 100. However, inspection of the raw data indicated that most participants detected and reset at least 60% of the offsets.

To test this hypothesis, the detection accuracy was scores were evaluated with the main effects test for complexity of the 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA. This main effect was significant, F (1, 23) = 8.32, p = .008; partial η2 = .27, a medium effect. Participants performed significantly better in the simple (*M* = 87.80, *SD* = 9.65) compared to the complex (*M* = 80.06, *SD* = 16.55) communication task.

For Study 1, in complex communication conditions participants were about 80% accurate in detecting system malfunctions compared to a nearly 88% accuracy rate when the communications were simple.

Hypothesis 4. The onset of communications will lead to poorer detections of system malfunctions depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 4a. Detection of system malfunctions will be lowest when communications occur simultaneously with or closest to system malfunctions with the worst rates of detection occurring for complex communications (a communication complexity by timing interaction effect).

As with the response time data, the hypothesized order of the effect for the worst to the best percent of detections of malfunctions was the 0, -2, -5, and -8 seconds complex communications followed by 0, -2, -5, and -8 seconds simple communication conditions.

Percent of correct detections were expected to follow a linear pattern with the lowest detections occurring in the time blocks where the communication and system malfunction were the closest. Complex communications were expected to elicit the lowest percent of correct detections. Means, standard deviations, and standard errors for the percent of correct detections for each of eight time blocks are presented in Table 8.

Table 8

Means, standard deviations, and standard errors for accuracy of detections (in percent) when communications occur before system malfunctions

Timing Interval & Complexity	Mean	SD	SE
0 second complex	68.45	26.30	5.37
-2 second complex	84.52	24.89	5.08
-5 second complex	82.14	20.74	4.23
-8 second complex	85.12	16.03	3.27
0 second simple	88.09	16.67	3.40
-2 second simple	86.90	18.80	3.84
-5 second simple	88.69	16.30	3.66
-8 second simple	87.50	12.15	2.87

The main effect for complexity was reported above. The main effect of timing on response accuracy was not significant: F (3, 69) = 2.42, p = .07; partial η2 = .09, a small effect.

The interaction of communication type and timing on response accuracy was significant, F (3, 69) = 3.34, p = .02; partial η2 = .10, a small effect. The post hoc comparisons for the interactions indicated significant differences for the simple simultaneous (*M*= 88.09; *SD*=16.67), -2 second (*M*= 86.90; *SD*=18.80), -5 second (*M*= 88.69; *SD*=16.30), and – 8 second (*M*= 87.49; *SD*=12.15) conditions compared to the complex (*M*=68.45; *SD*=26.30) simultaneous communication condition. None of the simple conditions were significantly different from each other.

The detection accuracy for the simultaneous complex communication condition was about 20% lower than for the simultaneous simple condition (see Figure 9). Also, no differences were found among the simple conditions, which argues against an onset preemption effect. While the hypothesis was only minimally supported, it highlights important distinctions between the timings of simple and complex communications relative to a detection event.

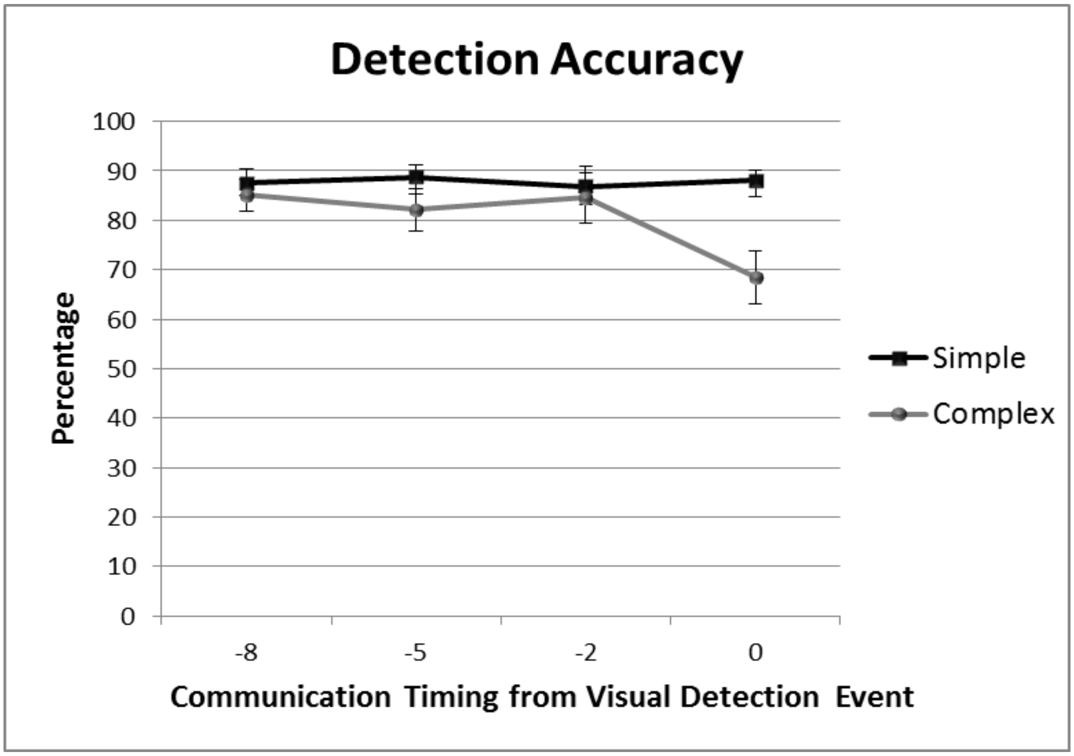

Figure 9. Detection accuracy (in percent) when communications occur before system malfunctions.

Hypothesis 4b. Detection of system malfunctions will not be impacted in conditions where either simple or complex communications occur at 8 seconds before the malfunctions (a null effect that tests the limits of echoic memory).

As with the response time data, this secondary hypothesis for percent of detections was assessed with a paired-samples t-test to compare the two, -8 second communication conditions (simple and complex). The results supported the hypothesis. There was not a statistically significant difference when comparing the -8 second simple (M = 87.50, SD = 12.15) and the -8 second complex (M = 85.12, SD = 16.03) communication conditions on the variable of percent of correct detections of system malfunctions, t (23) = .56, p = .58 (two-tailed), and a very small effect size, η^2= .01. The relationship between the two conditions can also be seen in Figure 9.

Together these findings for the response time to system malfunctions and the accurate detections of malfunctions indicate that in terms of auditory preemption theory, complex communications are more disruptive than simple communications, but only when they occur within the limits of echoic memory. Otherwise, participants performed similarly when simple and complex communications were presented at -5 or more seconds before a visual detection event.

In terms of magnitude of impact, on the response time measure, the difference between the least and most disrupted conditions was about 1.6 seconds (-2 second simple communications, M = 4.20 seconds vs. the simultaneous complex communication condition, M = 5.91 seconds). For the accuracy of detection measure, the least and most impacted conditions varied by over 20% (-5 second simple communication condition, M = 88.69 vs. the simultaneous complex communication conditions, M = 68.45).

The impact to the limitations of echoic memory appears to be strongly supported here. Communications were most disruptive to detection of malfunctions in an ongoing visual detection task when they occurred within the limits of echoic memory. However, it is complex communications when presented simultaneous to a malfunction that elicited the disruption to the ongoing visual task. Simple communications, even when presented simultaneously to a visual detection task, did not differ from each other in both the response time to and the accurate detection of malfunctions.

Beyond the limits of echoic memory (8 seconds or more), the operator may be able to resist the impact of communications on visual detections tasks under similar circumstances. It is within that boundary that strategic preemption or MRT principles regarding resource-sharing of auditory and visual tasks may apply, at least for these relatively straight-forward communication tasks.

Hypothesis 5. A concurrent, manual tracking task will be performed with more error in the presence of complex communications compared to simple communications (a simple effect).

The tracking task was measured as the root mean square error (RMSE) from the center of the tracking task window, expressed in pixel units. The data was recorded by the system at 1 second intervals. To test this hypothesis, the main effects test for complexity of the 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA was examined. The results did not support this hypothesis, and no significant effect was found, F (1, 23) = 0.23, p = .64; partial η2 = .01, a small effect. The mean RMSE was not statistically different when comparing the tracking task performance for the simple communication conditions (M = 29.53, SD = 6.73) to the tracking performance for the complex communication conditions (M = 29.90, SD = 7.88).

Hypothesis 6. The onset of communications will lead to more tracking task error depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 6a. Tracking error will be highest in conditions where communications occur simultaneously with or closest to system malfunctions with the highest tracking error occurring for complex communications (a communication complexity by timing interaction effect).

As with the other performance data, it was hypothesized that the tracking task RMSE data would indicate poorer performance when the communications task was paired with the system malfunctions in the order from 0, -2, -5, and -8 seconds complex communication conditions, followed by 0, -2, -5, and -8 seconds simple communication conditions. Means, standard deviations, and standard errors for the data are in Table 9.

Table 9

Means, standard deviations, and standard errors for tracking task RMSE when communications occur before system malfunctions

Timing Interval & Complexity	Mean	SD	SE
0 second complex	28.85	7.21	1.47
-2 second complex	29.23	8.13	1.66
-5 second complex	30.13	8.53	1.74
-8 second complex	31.38	9.10	1.86
0 second simple	29.25	6.40	1.31
-2 second simple	30.44	6.52	1.33
-5 second simple	29.61	8.11	1.66
-8 second simple	28.82	8.03	1.64

As reported above, the main effect for complexity was not significant. Also the main effect of timing on tracking task accuracy was not significant, F (3, 69) = 1.25, p = .30; partial $\eta 2$ = .05. Finally, the interaction of communication type and timing on tracking accuracy was not significant, F (3, 69) = 2.83, p = .06; partial $\eta 2$ = .11, a small effect. Figure 10 illustrates these results.

As discussed in the manipulation checks section earlier in the chapter, the RMSE data was collected at 1-second intervals and was summed and averaged across each of the time blocks rather than examined only for the few seconds of the communication/system malfunction pairings. It is possible that the noise in the data introduced by averaging all of the RMSE values across the block (including the off-pairing times) weakened the ability of statistical tests to find any differences. Future studies should examine the RMSE data with more fine-grained detail.

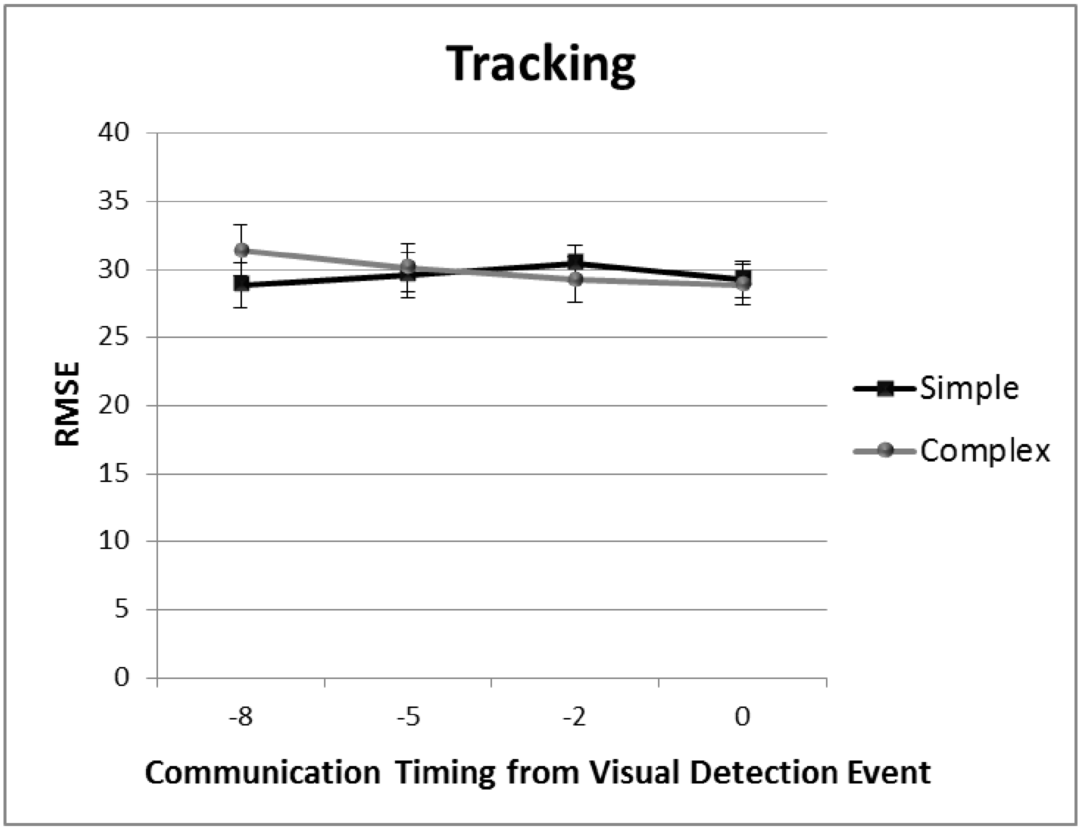

Figure 10. Tracking task error when communications occur before system malfunctions.

Hypothesis 6b. No differences will be observed for tracking task accuracy among the conditions where either simple or complex communications occur at 8 seconds before system malfunctions (a null effect that tests the limits of echoic memory).

A paired-samples t-test to compare the tracking task data for the two, -8 second conditions (simple and complex) was used to asses this secondary hypothesis for the tracking RMSE data. The results supported the hypothesis. There was not a statistically significant difference when comparing the -8 second simple (M = 28.83, SD = 8.03) and the -8 second complex (M = 31.38, SD = 9.10) conditions on the tracking RMSE variable, t (23) = -1.99, p = .06 (two-tailed), with a medium effect, η^2= .17. The relationship between the two conditions can be seen in Figure 10.

Hypothesis 7. Workload will be perceived as higher for complex communications compared to simple communications (a simple effect).

The NASA-TLX survey was collected after each timing block. The workload variable was the average of the six scales of the NASA-TLX. Each scale had a value from 0 to 100.

To test this hypothesis, the main effects test for complexity of the 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA was used. Results showed that the main effect of communication type on workload was significant, F (1, 23) = 18.19, p = .000, partial $\eta 2$ = .44. Participants reported lower workload for the simple (M = 38.23, SD = 18.52) compared to the complex (M = 44.62, SD = 18.14) communications tasks. Also, the difference between the ratings for the lowest and the highest rated conditions was nearly 9 points (8 second simple, M = 36.6 vs. simultaneous complex, M = 45.4).

Hypothesis 8. The onset of communications will result in higher subjective workload depending on the complexity of the communications as well as the limitations of human echoic memory.

Hypothesis 8a. Subjective workload will be highest when communications occur simultaneously with system malfunctions, followed by 2, 5 and 8 seconds before the malfunctions, with the highest workload occurring in the presence of complex communications at these timing intervals (a communication complexity by timing interaction effect).

It was hypothesized that the workload data would reflect the findings from the system monitoring response time data with the highest workload perceived for the 0, -2, -5, and -8 seconds complex communication conditions, respectively, followed by 0, -2, -5, and -8 seconds simple communication conditions. Table 10 contains the means, standard deviations, and standard errors for this data.

The main effect for complexity was significant, as reported above. However, the main effect of timing on workload was not significant: F (3, 69) = .63, p = .60; partial $\eta 2$ = .03, a small effect. And the interaction of communication complexity and timing on workload was not significant, F (3, 69) = .74, p = .53; partial $\eta 2$ = .03, also a small effect. Figure 11 illustrates the relationships. These results indicate that this hypothesis was partially supported.

Table 10

Means, standard deviations, and standard errors for workload when communications occur before system malfunctions

Timing Interval & Complexity	Mean	SD	SE
0 second complex	45.37	20.14	4.11
-2 second complex	44.32	18.24	3.72
-5 second complex	44.44	19.65	4.01
-8 second complex	44.34	18.23	3.72
0 second simple	37.71	19.08	3.89
-2 second simple	39.99	20.59	4.20
-5 second simple	38.64	18.75	3.83
-8 second simple	36.60	18.32	3.74

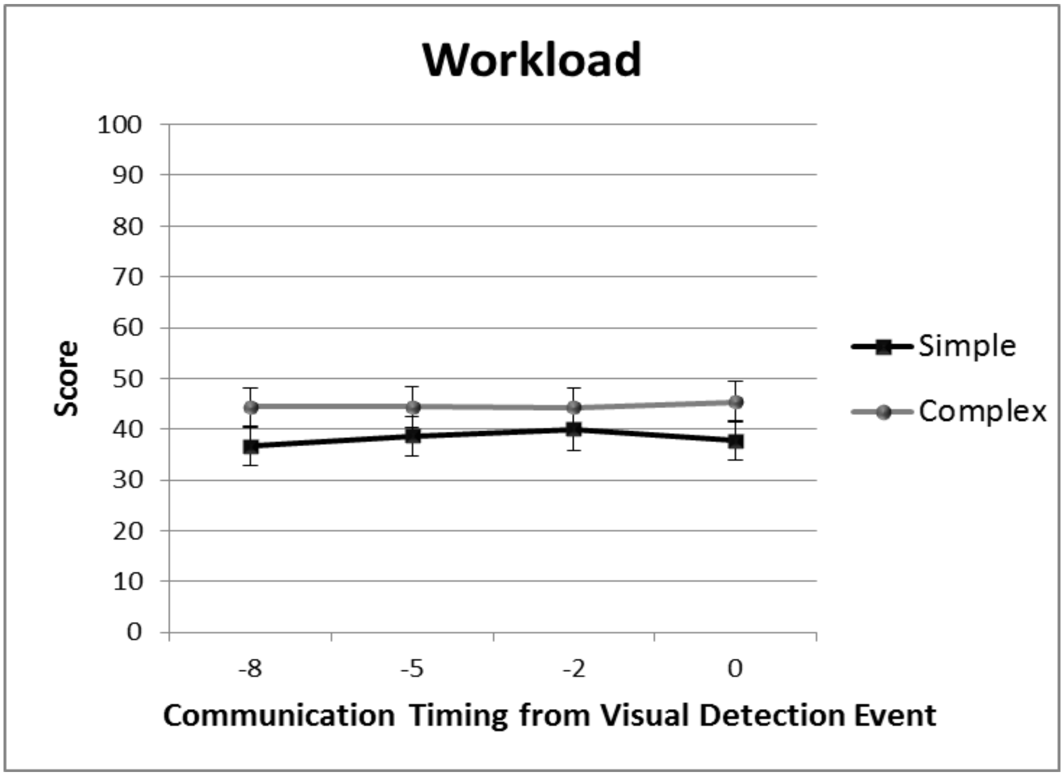

Figure 11. Subjective workload ratings when communications occur before system malfunctions.

Hypothesis 8b. No differences will be observed for subjective workload among the conditions where either simple or complex communications occur at 8 seconds before system malfunctions (a null effect that tests the subjective workload related to the limits of echoic memory).

A paired-samples t-test compared the workload data for the two, -8 second conditions (simple and complex) in order to asses this secondary hypothesis for subjective workload. The results did not support the hypothesis. A statistically significant difference was found when comparing the workload ratings for the conditions where communications were presented at 8 seconds prior to the visual detection task. For the simple communication condition, M = 36.60, SD = 18.32, and for the complex condition, M = 44.34, SD = 18.23, t (23) = -3.71, p = .001 (two-tailed). The effect size was medium, $\eta^2 = .37$. The relationship between the two conditions can be seen in Figure 11.

This result was somewhat surprising, especially considering that the t-tests on the performance data for the -8 second conditions (i.e. response time to system malfunctions and percent of detection malfunctions) were not significant. Nonetheless, participants did perceive workload as higher when complex communications were presented -8 seconds prior to a malfunction, compared to simple communications at the same timing interval. It is also possible that the higher subjective workload in the presence of no task performance differences mirrors prior studies regarding poor post-task estimates of performance (e.g., Lesch & Hancock, 2004).

Also, many participants commented after the study that they were surprised with the difficulty of the complex conditions and that their reflection about their past responses to complex communication requests had added an extra (and unanticipated) workload to the task. This finding indicates that a metacognitive process had been activated by the demands of the complex communication task and the subjective experience of workload may be the best indication of this assumption. What is unknown is whether the task demands activated the metacognitive process, or whether decisions to strategically preempt performance of the visual task while dealing with the auditory task activated metacognition. This observation presents intriguing insights for future research.

Study 2: Communication Occurring After System Malfunctions

Demographics, tests of random assignment, manipulation checks, and tests of normality were reported for both studies at the beginning of this chapter. They are not reported again here and the reader is referred to that earlier section for those analyses.

Study 2 Research Goal

The goal of Study 2 was to evaluate the impact of complexity and timing of communications when they occur <u>after</u> a malfunction in an ongoing visual detection task (i.e. the system monitoring task). This study was conducted to test alternative assumptions that some other factor, perhaps anticipation of communication events or auditory fatigue, would impact performance in the visual system monitoring task.

In order to match Study 1, the positioning of both simple and complex communication requests in Study 2 occurred at 2, 5 and 8 seconds <u>after</u> the malfunctions in the system monitoring task. However, since Study 1 included a condition that simultaneously presented a communication and system malfunction (i.e. the "0" second condition), Study 2 used a condition in which the separation of communications and malfunctions was well outside the limit of both human echoic and iconic memory stores. In that condition, 15 to 20 seconds occurred between system monitoring malfunctions for both simple and complex communications.

Since iconic memory lasts for only about 1 second, it was expected that participants would respond to the system malfunctions first before responding to the communication requests at these four timing intervals (i.e. 2, 5, 8 and 15 seconds after the visual detection of the system malfunction). Thus no interaction effects were anticipated. However, based on research that communications in general disrupt various types of transportation-related tasks, main effects were anticipated for simple versus complex communications on study dependent variables.

Tests of Hypotheses

Tests of the main hypotheses for Study 2 were conducted in the same fashion as Study 1, using a series of 2 (communication complexity: simple and complex) x 4 (communication timing: 0, 2, 5, and 8 seconds) factorial repeated measures ANOVAs. Several sub-hypotheses were evaluated using paired samples t-tests with 2-tailed tests of significance. Estimates of effect sized are based on those provided in the G*Power program (Faul, Erdfelder, Lang, & Buchner, 2007) as well as in Pallant (2007) and Cohen (1988). Graphics accompany the hypotheses, as appropriate.

Hypothesis 1. Response time to system malfunctions will be longer in conditions with complex communications compared to simple communications conditions, and no interaction of complexity and timing is anticipated (a simple effect).

To evaluate this hypothesis, the data for each complexity by timing condition was submitted to a 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA. Results indicated no main or interaction effects for detection response times ($p>.05$). Means, standard deviations, and standard errors for the data are in Table 11. The relationships are shown in Figure 12.

These two tests indicated that this hypothesis was not supported. Response times to system malfunctions did not vary by complexity of communications that occurred just after the malfunction. There also was no effect for the timing of communication requests.

Table 11

Means, standard deviations, and standard errors for response times when system malfunctions occur before communications

Timing Interval & Complexity	Mean	SD	SE
2 second complex	4.22	1.32	0.27
5 second complex	4.15	1.39	0.28
8 second complex	3.99	1.97	0.40
15+ second complex	4.01	1.37	0.28
2 second simple	3.87	1.54	0.32
5 second simple	3.72	1.65	0.34
8 second simple	3.82	2.22	0.45
15+ second simple	4.02	1.31	0.27

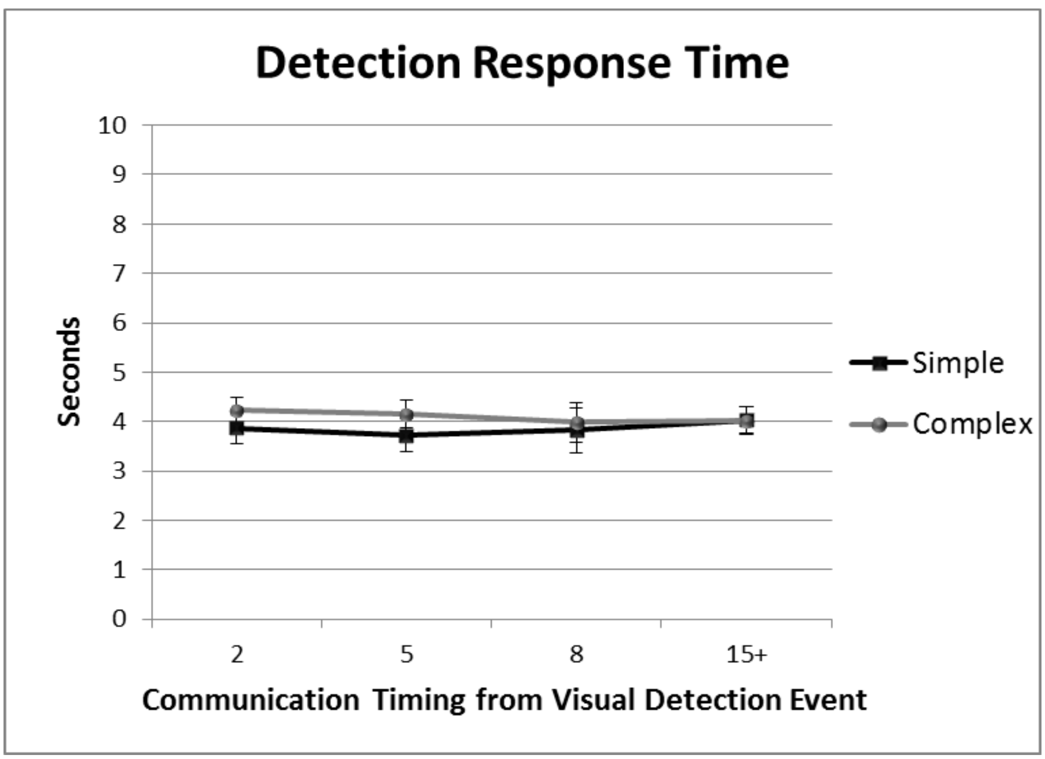

Figure 12. Response times when malfunctions occur before communications.

Hypothesis 2. Percent of correct detections of system malfunctions will be lower for complex communications compared to simple communications conditions, and no interaction of complexity and timing of communications is anticipated (a simple effect).

The data for each condition (see

Table 12) was submitted to a 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA to evaluate any communication complexity by timing effects for the percent of correct detections variable. Results were not significant for main and interaction effects. Percent of correct detections did not vary in Study 2 by complexity of the communications. There also was no effect for the timing of the presentation of the communications task when it occurred after the malfunction. Results are shown in Figure 13.

Table 12

Means, standard deviations, and standard errors for accuracy of detections (in percent) when system malfunctions occur before communications

Timing Interval & Complexity	Mean	SD	SE
2 second complex	91.67	11.08	2.26
5 second complex	91.07	11.00	2.24
8 second complex	88.69	18.35	3.75
15+ second complex	91.07	14.48	2.96
2 second simple	92.26	13.96	2.85
5 second simple	92.86	11.92	2.43
8 second simple	90.48	20.06	4.09
15+ second simple	92.26	9.40	1.92

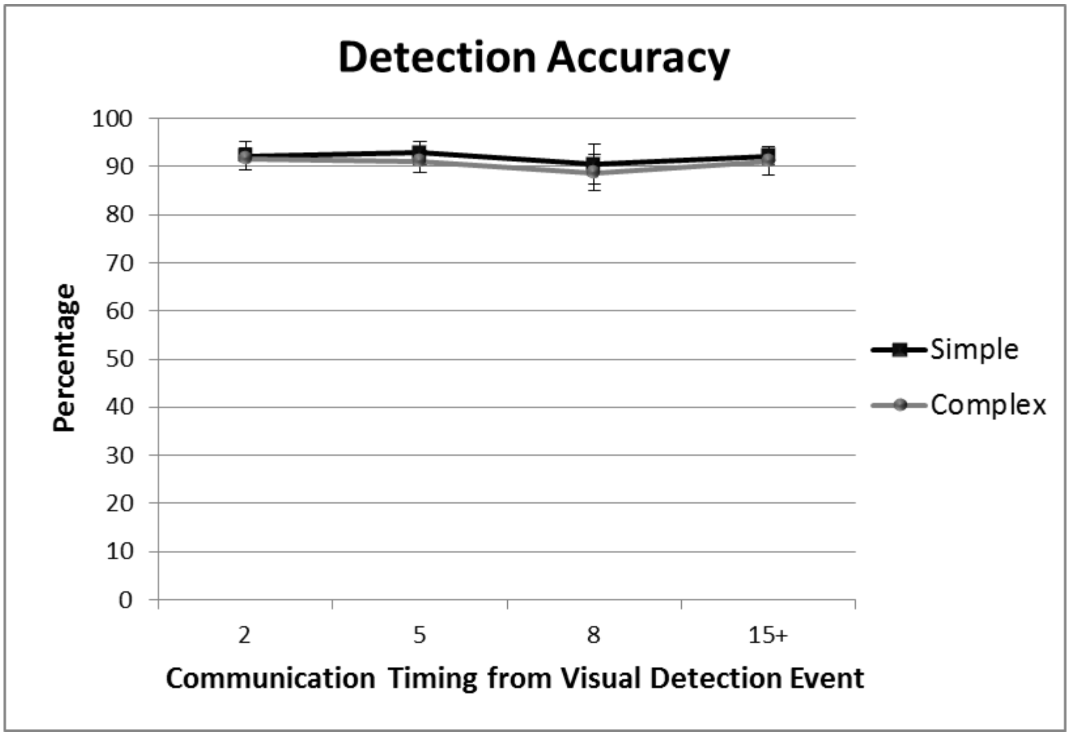

Figure 13. Detection accuracy (in percent) when malfunctions occur before communications.

Hypothesis 3. A concurrent, manual tracking task will be performed with more error in the presence of complex communications compared to simple communications, and no interaction of complexity and timing of communications is anticipated (a simple effect).

The participants' RMSE data for the eight complexity and timing blocks were also evaluated with 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA. The means, standard deviations, and standard errors are shown in Table 13. Figure 14 shows this analysis.

Results were not significant for main and interaction effects ($p > .05$). This hypothesis for an effect on tracking performance accuracy of complexity of communications when they occur close in time but just after a monitoring malfunction was not supported. There was no effect for complexity or timing of this data.

Table 13

Means, standard deviations, and standard errors for tracking RMSE when system malfunctions occur before communications

Timing Interval & Complexity	Mean	SD	SE
2 second complex	29.09	8.75	1.50
5 second complex	28.90	7.33	1.46
8 second complex	28.97	7.15	1.59
15+ second complex	29.22	7.80	1.79
2 second simple	28.86	7.33	1.46
5 second simple	29.05	7.14	1.64
8 second simple	29.31	8.03	1.56
15+ second simple	29.63	7.63	1.50

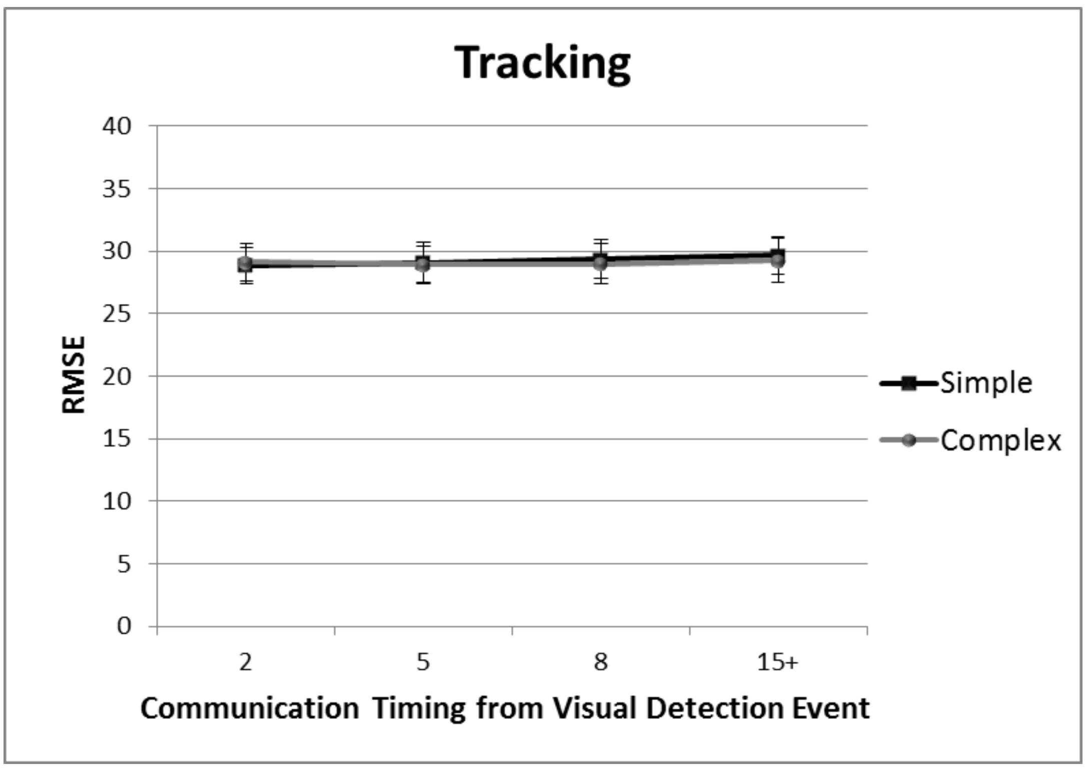

Figure 14. Tracking task error when malfunctions occur before communications.

Hypothesis 4. Workload will be perceived as higher for complex compared to simple communication conditions, and no interaction of complexity and timing of communications is anticipated (a simple effect).

The workload data for study 2 was analyzed with a 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA. Table 14 contains the means, standard deviations, and standard errors for the Study 2 workload data.

Results indicated that the min effect of complexity type on workload was significant, F (1, 23) = 10.17, p = .004; partial η2 = .31, a medium to large effect. Participants reported lower workload for the simple (M = 46.27, SD = 15.19) compared to the complex (M = 51.88, SD = 12.93) communications tasks. Also the main effect of timing on workload was not significant, F (3, 69) = .65, p = .59; partial η2 = .03, a small effect. And the test for interaction of complexity type and timing on workload was not significant, F (3, 69) = .59, p = .63; partial η2 = .02. The relationships are shown in Figure 15.

Table 14

Means, standard deviations, and standard errors for workload when system malfunctions occur before communications

Timing Interval & Complexity	Mean	SD	SE
2 second complex	53.69	13.96	2.85
5 second complex	51.61	13.94	2.85
8 second complex	50.15	14.69	3.00
15+ second complex	52.06	13.74	2.81
2 second simple	46.10	16.15	3.30
5 second simple	46.24	15.28	3.12
8 second simple	46.45	15.98	3.26
15+ second simple	46.28	17.44	3.56

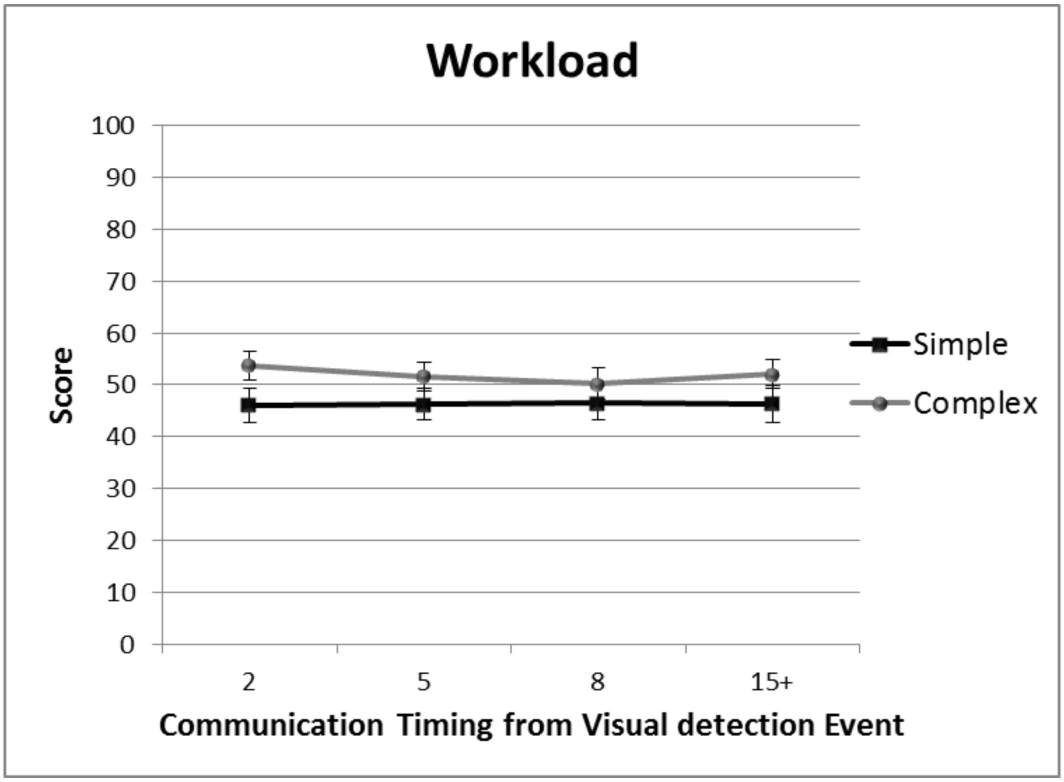

Figure 15. Subjective workload when malfunctions occur before communications.

A comparison of the means for the Study 1 versus the Study 2 workload data revealed that participants rated workload as higher in Study 2 for every comparable condition. All of the simple communications in Study 2 were rated between 7 and 10 points higher than the Study 1 simple conditions and all of the complex conditions were rated as higher by 8 to 9 points in Study 2 versus Study 1. It was noted that in all of the conditions in Study 2 the workload ratings were more variable than the Study 1 group. The conditions in Study 2 ranged between 3 and 7 points higher in standard deviation than those in Study 1.

Additional Analyses

Individual Differences

Researchers in the area of human factors and ergonomics have called for an increased emphasis and attention to be paid to individual differences in performance in both research and practice (Szalma, 2009). One individual difference is that of performance between males and females on various tasks.

A construct closely related to the overall tasks in this study is complacency, or an operator's tendency to miss critical events in a system when monitoring occurs over an extended period of time (Mouloua, Parasuraman, & Molloy, 1993; Parasuraman, Mouloua, & Molloy, 1996; for a review, see Mouloua, Hancock, Jones & Vincenzi, 2010). The performance consequences of these monitoring failures have included missing critical events and delayed reaction times especially under a time pressure (Parasuraman & Manzey, 2010). Performance on secondary tasks has also been known to suffer under increased workload in detection tasks (Vincenzi & Mouloua, 1998). With regard to sex differences in complacency, Parasuraman and Manzey (2010) indicated that no differences appear to exist in these studies, but the authors also note that the available studies related to individual differences is small and decisive conclusions are not advised. Therefore, an examination of differences between males and females on performance measures in the current research is warranted.

One consistent finding regarding sex differences has been that males perform better on visual-spatial or spatial ability tasks than females starting by early adolescence and continuing through adulthood (Maccoby & Jacklin, 1974, p. 91). In fact in their comprehensive review of studies in the area to that time, those researchers found a performance difference of at least .40 standard score units in favor of males by the end of high school.

However in the area of perceptual motor abilities, Maccoby and Jacklin's (1974, p. 38) summary of research indicated that by high school, boys show greater speed and coordination of gross body movements. On the other hand, females excel in tasks of visual-motor coordination, finger dexterity, and motor coordination, with differences in those tasks ranging from 5 percent to 10 percent higher than males. The sexes have not been found to consistently differ in manual dexterity, and the authors note the importance of defining tasks that require large muscle versus fine muscle movements.

In other research, cognitive differences have been shown with regard to the way that males and females perceive time. Hancock (2011) summarized the history of this research and then synthesized 30 time perception studies, plotted the studies on a graph by males and females, and provided the results of a linear regression equation from the data. On the graph, the duration of the perception of judgment across the studies was 120 seconds and was plotted against the actual target time of 120 seconds. The slope value for the regression for female participants was 1.13 versus a slope value of nearly 1 for males, or a 13% longer temporal estimation of time by females (p. 189).

These average results over the several studies indicated that males are better able to estimate time clock time than females. In his introduction, Hancock (2011) summarized the findings this way: "…there are consistent temporal processing differences between the sexes and these differences are of the same order of magnitude as the already known spatial processing differences (p. vii)." Hancock (2011) also emphasized that the only reliable sex difference in other areas of performance has been with regard to spatial abilities, and considering space that and time are fundamentally related, the further study of sex differences related to time should not continue to be neglected.

In the current research, the compensatory tracking task combines aspects of visual-spatial ability, manual outputs, and quick performance and presents the best opportunity to assess sex differences. When using the root mean square error, which combines both directionality (arguably a spatial ability factor) and time to return the target to center (a velocity, speed or time factor), it is possible that the lack of findings presented earlier for the tracking task in the current research may have been masked by individual differences in performance of the task.

With regard to tracking task performance, prior research has indicated variability between relatively unskilled versus skilled performers (Miyake, Loslever, & Hancock, 2001) such that poorer performers tend to use an up-vertical and low-velocity combination of actions while skilled individuals use all directions and high velocity inputs in the task. However, only males were used in that study so no sex differences were evaluated.

Based on the review above, the current research offers the opportunity to contribute to human factors knowledge regarding individual differences. To accomplish these analyses, a series of independent samples t-tests comparing males' and females' scores by timing blocks on each of the variables in the two studies were performed. For variables where the t-tests were significant, an additional 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA was performed using sex (male/female) as a covariate. The results for each variable for the two studies are reported together.

Response time.

As shown in Table 15, none of the independent samples t-tests evaluating males' and females' response times to system malfunctions were significant. Males and females in this study did not differ in their response times to the visual detection events when communications occurred at the various timing intervals before or after those events. No further analyses were performed on this measure.

Table 15

Means, standard deviations, and t-tests comparing males' and females' response times to system malfunctions by study and timing blocks

		Simple				sig.		Complex				sig.	
Timing Blocks	Sex	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*
Study 1													
- 8 seconds	Male	4.96	*1.03*	*1.07*	*22*	.30	.05	4.30	*1.86*	-.40	*22*	.70	.01
	Female	4.41	*1.43*					4.57	*1.49*				
- 5 Seconds	Male	4.67	*2.21*	*.94*	*22*	.36	.04	5.00	*2.49*	.22	*22*	.83	.002
	Female	3.97	*1.41*					4.82	*1.50*				
- 2 Seconds	Male	4.39	*2.01*	*.49*	*22*	.63	.01	5.48	*1.65*	.30	*22*	.77	.004
	Female	4.04	*1.50*					5.25	*2.14*				
Simultaneous	Male	5.21	*2.20*	*1.05*	*22*	.30	.05	6.21	*2.35*	.64	*22*	.53	.02
	Female	4.53	*0.72*					5.66	*1.86*				
Study 2													
2 seconds	Male	3.49	1.85	-.92	22	.37	.04	4.13	*1.09*	-.25	*22*	.80	.003
	Female	4.09	1.35					4.28	*1.48*				
5 seconds	Male	3.47	1.57	-.57	22	.58	.01	3.86	*1.27*	-.80	*22*	.43	.03
	Female	3.87	1.73					4.33	*1.47*				
8 seconds	Male	3.45	2.33	-.64	22	.53	.02	3.23	*1.64*	-1.49	*22*	.15	.09
	Female	4.05	2.21					4.44	*2.06*				
15+ seconds	Male	3.52	1.26	-1.51	22	.15	.09	3.81	*1.66*	-.54	*22*	.59	.01
	Female	4.33	1.29					4.13	*1.21*				

*Significant effect

Accuracy of detections.

The t-tests for the accuracy of detections of system malfunctions also were not significant. The results in Table 16 show that males and females did not differ in their accuracy of detecting system malfunctions considering both the timing and the complexity of an interrupting communication task. No further analyses were performed on this measure.

Table 16

Means, standard deviations, and t-tests comparing males' and females' detection accuracy by study and timing blocks

		Simple				sig.		Complex				sig.	
Timing Blocks	Sex	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*
Study 1													
- 8 seconds	Male	85.71	*11.07*	-.65	*22*	.18	.02	84.42	*18.58*	-.59	*22*	.56	.02
	Female	89.01	*13.24*					85.71	*14.29*				
- 5 Seconds	Male	87.01	*21.62*	-.46	*22*	.38	.01	75.32	*27.17*	.21	*22*	.84	.002
	Female	90.11	*10.73*					87.91	*11.44*				
- 2 Seconds	Male	83.11	*21.96*	-.90	*22*	.65	.03	85.71	*23.91*	-1.52	*22*	.14	.09
	Female	90.11	*15.85*					83.52	*26.63*				
Simultaneous	Male	83.12	*21.96*	-1.37	*22*	.52	.08	64.93	*28.83*	-.19	*22*	.85	.002
	Female	92.31	*9.43*					71.43	*24.74*				
Study 2													
2 seconds	Male	90.48	*18.90*	-.48	*22*	*.64*	.01	88.89	*11.90*	-.95	*22*	.35	.04
	Female	93.33	*10.62*					93.33	*10.62*				
5 seconds	Male	92.06	*14.48*	-.25	*22*	*.81*	.003	92.06	*10.38*	.34	*22*	.74	.005
	Female	93.33	*10.62*					90.48	*11.66*				
8 seconds	Male	87.30	*28.07*	-.59	*22*	*.56*	.02	90.47	*10.10*	.36	*22*	.72	.006
	Female	92.38	*14.15*					87.62	*22.17*				
15+ seconds	Male	96.82	*6.30*	1.95	*22*	*.06*	.14	92.06	*14.48*	.25	*22*	.80	.003
	Female	89.52	10.05					90.47	*14.95*				

*Significant effect

Tracking.

The t-tests comparing males and females performance on the tracking task found no significant results in the simple communication conditions, as shown in Table 17. Males and females did not differ on tracking performance at any of the timing intervals when they were required to repeat words that they heard either before or after a visual detection event. In addition, performance did not differ between males and females when they were required to generate three words in the complex communication conditions that occurred at timing intervals just after visual detection events (i.e. study 2).

However, when the task required generating three words at timing intervals just before the visual detections events (i.e. study 1), sex differences were found. Specifically, the t-tests indicated that males (*M* = 27.25, *SD* = 7.98) performed with significantly less tracking error than females (*M* = 34.87, *SD* = 8.78) when they were required to generate three words at -8 seconds prior to the visual detection event; *t* (22) = -2.21, *p* = .04 (two-tailed), a small to medium effect size, η^2= .17. In addition, males (*M* = 26.10, *SD* = 6.80) performed with significantly less tracking error than females (*M* = 33.54, *SD* = 8.56) when required to respond to the complex communications that occurred -5 seconds prior to the visual malfunctions; *t* (22) = -2.33, *p* = .03 (two-tailed), a small to medium effect, η^2= .19.

Table 17

Means, standard deviations, and t-tests comparing males' and females' tracking error by study and timing blocks

		Simple				sig.		Complex				sig.	
Timing Blocks	Sex	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*
Study 1													
- 8 seconds	Male	28.22	*9.66*	-*.33*	*22*	.74	.005	27.25	*7.98*	-2.21	*22*	.04*	.17
	Female	29.34	*6.74*					34.87	*8.78*				
- 5 Seconds	Male	28.38	*9.20*	-*.67*	*22*	.51	.02	26.10	*6.80*	-2.33	*22*	.03*	.19
	Female	30.64	*7.28*					33.54	*8.56*				
- 2 Seconds	Male	27.95	*4.57*	*-1.80*	*22*	.09	.12	26.66	*7.28*	-1.46	*22*	.16	.08
	Female	32.55	*7.33*					31.40	*8.44*				
Simultaneous	Male	27.69	*5.66*	*-1.10*	*22*	.28	.05	26.74	*7.60*	-1.34	*22*	.19	.07
	Female	30.57	*6.91*					30.63	*6.64*				
Study 2													
2 seconds	Male	28.63	6.30	-.12	22	.90	.001	28.45	*5.84*	-.33	*22*	.74	.005
	Female	29.00	7.81					29.48	*8.26*				
5 seconds	Male	29.33	8.88	.09	22	.93	.001	29.09	*7.45*	.09	*22*	.93	.001
	Female	28.88	7.79					28.80	*7.23*				
8 seconds	Male	29.29	7.96	.00	22	1.00	.001	28.98	*7.02*	.00	*22*	1.00	.001
	Female	29.32	7.71					28.96	*8.48*				
15+ seconds	Male	28.33	4.69	-.28	22	.78	.003	28.56	*8.06*	-.28	*22*	.78	.003
	Female	30.40	8.61					29.62	*9.39*				

*Significant effect

To further evaluate this sex difference, the tracking data for study 1 was submitted to a 2 (communication complexity type) x 4 (communication timing) factorial repeated measures ANOVA using sex as a covariate. That analysis indicated no significant main effect for communication type, *F* (1, 22) = 4.03, p = .06; partial η2 = .16, or for timing, *F* (3, 66) =.25, p = .87; partial η2 = .01. There was a significant interaction between communication type and sex, *F*

(1, 22) =5.18, $p = .03$; partial $\eta 2 = .19$, a small to medium effect, but not between timing and sex, F (3, 66) =.62, $p = .61$; partial $\eta 2 = .03$. The test for the three way interaction between type, timing and sex was significant, F (3, 66) =2.78, $p = .05$; partial $\eta 2 = .11$, a small effect.

Post hoc tests indicated that significantly more tracking task error occurred in the -8 second complex condition than in the -2 second or simultaneous complex conditions. In addition, the difference between the -8 second complex condition for females was significantly different than the -8 second simple condition for females and the -8 second simple and complex conditions for males. The post hoc tests did not find significant differences among the -5 second simple or complex conditions when sex was used as a covariate in the factorial repeated measures analysis. Figure 16 shows the effect.

An additional one-way repeated measures ANOVA was performed using only the females' tracking scores for four complex communications conditions, as shown in Figure 16, to evaluate the apparent trend in this data. The results were significant, F (3, 36) =5.32, $p = .004$; partial $\eta 2 = .31$, a medium to large effect. The post hoc analysis indicated that the -8 second condition was significantly higher than the -2 and simultaneous conditions, and the -5 second condition was significantly higher than the simultaneous condition. In this study, the timing of the complex communications in relation to the visual detection events was an important factor in females' performance on the tracking task. A comparable analysis was performed with the four simple communication conditions for females, but the effect was not significant, F (3, 36) =2.70, $p = .06$; partial $\eta 2 = .18$, a small effect.

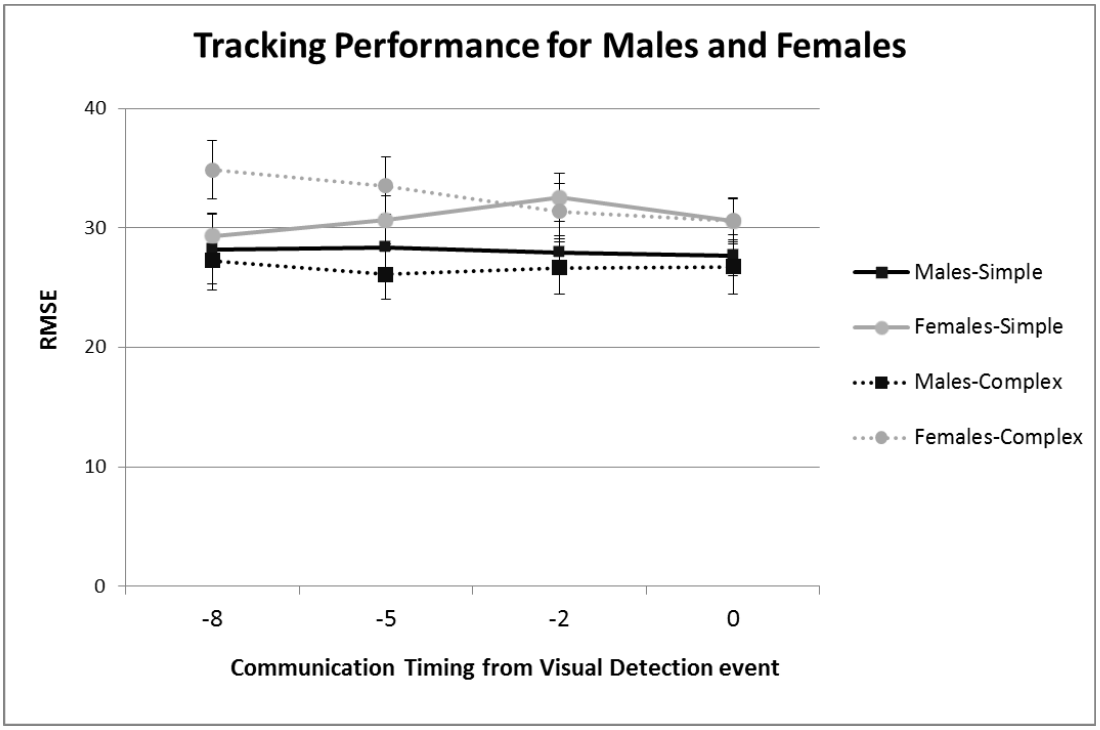

Figure 16. Tracking task error for males and females when communications occur before system malfunctions.

Overall, when sex is taken into account it was the females' performance in the -8 second complex communication condition, with a linear trend for better performance as communications occurred closer to a visual detection event, which accounted for individual differences in the tracking task. These results indicate that the additional task loading of communications may actually improve female performance for a visual-spatial task in a multitask system. For males, this trend was not noted. Their tracking performance did not appear to change significantly in the presence of communication tasks. This variable trend in performance for females may indicate, as Hancock (2011) explains, that space and time are fundamentally intertwined and that prior research regarding sex differences in spatial tasks may also extend to similar differences for temporal tasks. Further research regarding these interactions is warranted.

Subjective workload.

The t-tests comparing males' and females' performance on workload found no significant results for the simple or complex communication conditions when they occurred before visual detection events, or for simple communications when they occurred after a detection event (see Table 18). However, for the complex communications when they occurred after malfunctions (study 2), there was one significant t-test for the 8 second condition. Females (M = 55.23, SD = 14.99) rated their workload as significantly higher at that timing than males (M = 41.67, SD = 9.89), t (22) = -2.41, p = .02 (two-tailed), a small to medium effect, η^2= .20.

To further evaluate this finding, the study 2 workload data was analyzed in a 2 (communication complexity) x 4 (communication timing) factorial repeated measures ANOVA using sex as a covariate. The results indicated no main effects for communication type, F (1, 22) = .42, p = .52; partial η2 = .02; or for timing, F (3, 66) =1.15, p = .34; partial η2 = .05; or for the interaction of communication type and sex, F (1, 22) =.06, p = .80; partial η2 = .003; or for the interaction of timing and sex, F (3, 66) =.95, p = .42; partial η2 = .04; or for the three way interaction of timing, complexity, and sex, F (3, 66) =1.66, p = .18; partial η2 = .07. Entering sex as a covariate into the repeated measures ANOVA did not improve on the model. The effect of sex on workload ratings does not appear to be a robust one.

Table 18

Means, standard deviations, and t-tests comparing males' and females' workload by study and timing blocks

		Simple				sig.		Complex				sig.	
Timing Blocks	Sex	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*	*M*	*SD*	*t*	*df*	*(2-tailed)*	*η2*
Study 1													
- 8 seconds	Male	37.01	*16.78*	.10	*22*	.92	.03	41.49	*20.07*	-.70	*22*	.49	.02
	Female	36.24	*20.21*					46.75	*16.95*				
- 5 Seconds	Male	36.62	*18.01*	-*.48*	*22*	.64	.01	39.09	*21.29*	-1.24	*22*	.23	.06
	Female	40.35	*19.92*					48.97	*17.72*				
- 2 Seconds	Male	37.82	*18.82*	-*.47*	*22*	.64	.009	43.49	*20.08*	-.20	*22*	.84	.00
	Female	41.83	*22.57*					45.03	*17.32*				
Simultaneous	Male	33.92	*18.39*	-*.89*	*22*	.38	.03	44.62	*22.37*	-.16	*22*	.87	.00
	Female	40.91	*19.79*					46.00	*18.97*				
Study 2													
2 seconds	Male	41.48	*19.05*	-1.09	*22*	.29	.05	49.28	*13.95*	-1.21	*22*	.24	.06
	Female	48.87	*14.12*					56.34	*13.73*				
5 seconds	Male	42.19	*18.41*	-1.01	*22*	.33	.04	46.19	*13.34*	-1.52	*22*	.14	.09
	Female	48.67	*13.15*					54.87	*13.68*				
8 seconds	Male	42.69	*19.14*	-.89	*22*	.38	.03	41.67	*9.89*	-2.41	*22*	.02*	.20
	Female	48.71	*13.99*					55.23	*14.99*				
15+ seconds	Male	37.87	*16.56*	-1.94	*22*	.07	.14	47.20	*12.92*	-1.36	*22*	.19	.07
	Female	51.32	*16.43*					54.97	*13.81*				

*Significant effect

Evaluation of Multitask Environments

This research has had the overarching goal to evaluate the impact of the complexity and timing of communication interruptions in relation to operator performance in a multitasking environment. Many psychological studies of operator performance under multitasking are actually dual-task studies where the impact of one task is evaluated as the outcome in performance on a second task. Implications for variation in performance are then discussed in terms of psychological theories or projected impacts on the multitask environment.

One significant limitation for multitask studies using psychological measures is that there is rarely one common scaling for comparison across outcome variables making it difficult to equate performance on more than two tasks. For example, in the current study, task accuracy was measured on a percent of correct detections scale with 100 being a perfect score. Response time was measured in seconds, and the longest time recorded was up to 10 seconds with any response after that counted as a miss and valued at 10 seconds. Tracking was measured as root mean square error from center in terms of screen pixel units with a range of 0 (exact center) to about 60 (the most outlying scores obtained). Finally workload was measured on a 0 to 100 scale.

In addition to common scaling, the direction of scaling of each variable should be considered. Of the four measures in this study, the accuracy scale equates higher numbers to better performance. The other three scales (i.e. response time, tracking, and subjective workload) link higher numbers to worse performance. It is these complexities in evaluation that make multitask comparisons difficult.

In fields such as engineering design and project management, an approach to evaluating the performance or effectiveness of a system is to establish a figure of merit (FOM; see Lee, 2011, for a brief explanation). Such an approach is similar to performance metrics or measurements used in various areas of business, accounting, marketing and organizational management (e.g. Neely, 2007). An FOM is a number, figure, or other criterion that is used to estimate the efficiency, usefulness, or other attribute of a design, system, process, or product, often in relation to alternatives. These are often expressed as ratios. For example, from consumer products common FOMs are the miles per gallon ratings for cars, resolution of television screens, or megapixels of a camera. Sports statistics such as earned run average, free throw percentage, and the handicap rating in golf may also be considered figures of merit. While an FOM is not commonly reported in research studies regarding human user performance in systems, examples and discussions of applications of this technique are available (e.g. LeMay & Comstock, 1990; Hartson, Andre & Williges, 2003; Chattratichart & Lindgaard, 2009).

Using a figure of merit approach for the current data, z-scores were calculated for each individual's score on each measure using the grand means and standard deviations across all conditions (i.e. all timing blocks) in both studies for the specific measure. For example, each individual's detection accuracy score was subtracted from the grand mean of all participants' accuracy scores from both studies and divided by the standard deviation of all participants' accuracy scores in both studies. After this standardization of scores, the means, standard deviations and standard errors were computed for each timing condition. The standard deviation and standard error terms by condition were used for t-tests and for error estimates for the means in the graphs presented, similar to other psychological analyses. The complete table of mean z-scores, standard errors, and standard deviations by measures and conditions for both simple and complex communications can be found in Appendix G.

These z-scores with their standard errors for each measure by complexity level were then plotted on graphs for examination. It should be noted that in the graphs, the performance measures z-score axis was truncated as -2 through +2 standard deviation units to aid visualization of the relative differences among the conditions.

Figure 17 shows the relative position of participants' performance on the system measures (i.e. accuracy of detections, response time to the visual detection events, and tracking) and the associated workload ratings for each of the eight timing blocks for simple communications. Examination of this figure shows potential performance trends across timing intervals for simple communications. For example, the lines start to diverge at about -2 seconds and then spread until about +2 seconds, a trend noted from the results sections reported earlier. This apparent perturbation across the timing intervals indicates variable performance on system tasks when a communication is presented close to a visual detection event.

Also after the 2 second interval, the lines do not intertwine as they had at the -8 and -5 timing intervals. Further of note, the response time measure improves (i.e. drops) and response accuracy improves (i.e. rises) in the intervals at 2 seconds and higher, while the tracking measure appears to hold a central position relative to the other measures across all intervals. It is possible that these performance measures can provide a different look at human performance depending on the collocation of a communication with a visual detection event.

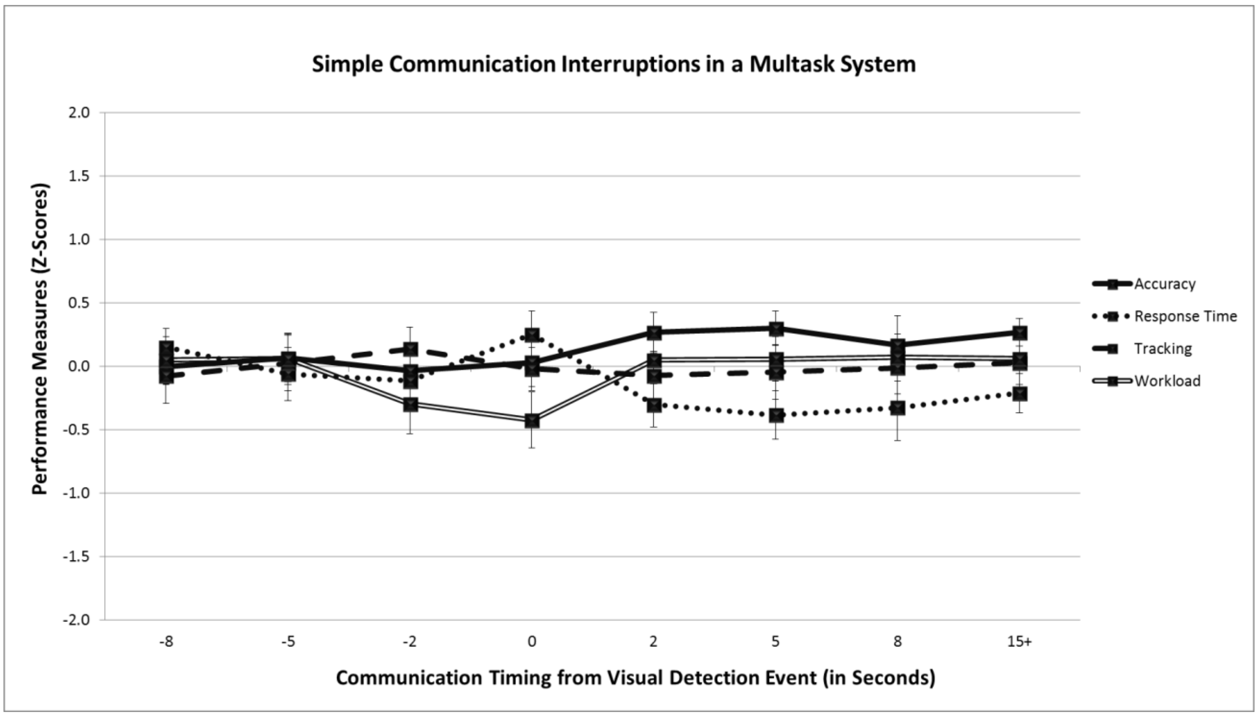

Figure 17. Z-scores and standard errors for performance measures for simple communications across timing intervals.

Similarly, Figure 18 shows the relative position of participants' performance and associated workload ratings for complex communications. Examination of this figure shows essentially the same trends as in the simple communications, although much more variability across the scores is evident. Of note is a linear trend for increased response time as complex communications are presented closer, but prior to, the visual detection events. This also was discussed in the earlier results section. The same perturbation from -2 to +2 seconds as in simple communications is present, and the effects for the drop in accuracy and the increase in response time are obviously significant at the simultaneous timing interval. Compared to the simple communications graph, these performance measures during complex communications offer different perspectives on performance across the timing blocks.

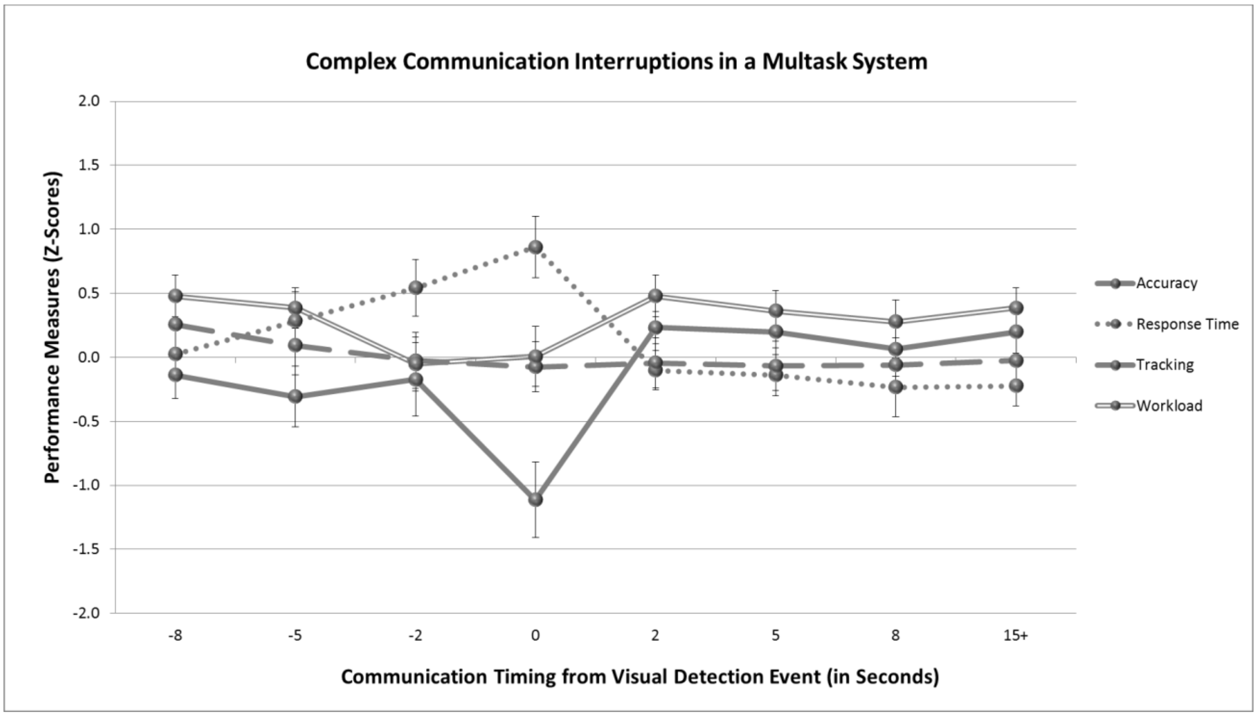

Figure 18. Z-scores and standard errors for performance measures for complex communications across timing intervals.

For more direct comparison of the measures for both simple and complex communications across the timing intervals, all of the performance measures were plotted on one graph, as seen in Figure 19. Comparison of the positions of the simple communication performance measures (black lines) with the complex performance measures (grey lines) show both the variability of performance relative to each other, as well as the variability of the scores relative to a mean z-score (represented as the central, 0.0 line). Again, the differences across the timing intervals indicates that decisions about the effectiveness of a system may be substantially influence by the choice of measures for figures of merit, an argument that has been made by other authors (e.g., Watjatrakul, 2005).

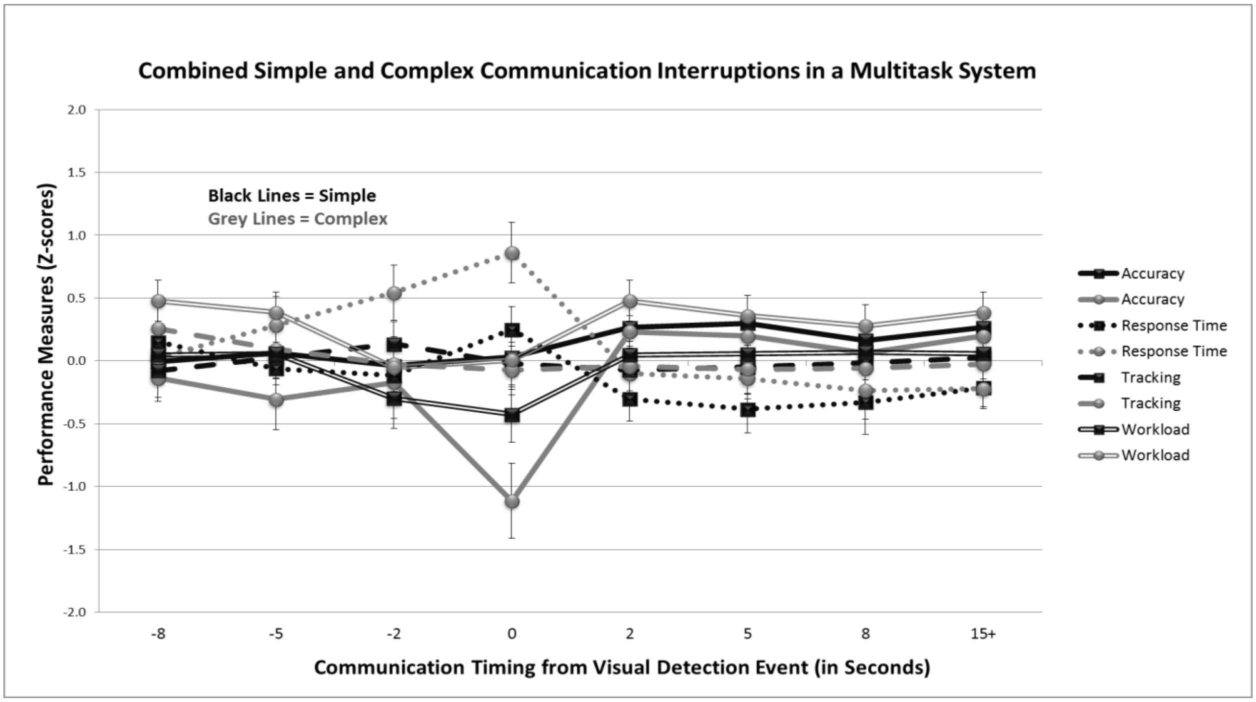

Figure 19. Z-scores and standard errors for performance measures for combined simple and complex communications across timing intervals.

Following the plots of the measures shown above, the figures of merit were computed for each of the eight timing intervals by communication complexity. Since higher accuracy ratings were equated with better performance, but higher scores on the other three performance measures were equated with worse performance, the accuracy measure z-scores were inverted so that the scaling for all four measures were in the same direction. This was accomplished simply by changing the signs of the z-score means for each condition for the accuracy scores only.

In some uses of figures of merit, weights are assigned to various measures depending on the theorized or actual importance of a measure to the overall effectiveness of a system.

However, since there was no theoretical reason to assign weights to any of the performance measures for this study, the figures of merit here were derived by simply averaging values.

Therefore, 16 figures of merit were computed by averaging the z-score means for each of the four performance score values for each timing interval at each level of complexity. The same averaging method was used for standard errors and standard deviations. These average values are shown in Table 19. The pairs of FOMs for each timing interval were further analyzed with a series of eight t-tests to compare the simple versus the complex conditions. These t-tests and effect sizes also are shown in Table 19. Recall that negative values mean better performance.

Table 19

Figures of merit data for timing intervals, communication complexity levels, and t-test comparisons

Timing Intervals	Complexity	*M*	*SD*	*SE*	*t*	*df*	sig. *(2-tailed)*	*η2*
- 8 seconds	Simple	0.03	0.84	0.17	.72	46	.47	.02
	Complex	0.22	0.96	0.20				
- 5 Seconds	Simple	-0.01	1.00	0.20	.97	46	.34	.04
	Complex	0.27	1.05	0.21				
- 2 Seconds	Simple	-0.06	1.02	0.21	.71	46	.48	.02
	Complex	0.16	1.15	0.23				
Simultaneous	Simple	-0.06	0.94	0.19	1.76	46	.08	.12
	Complex	0.48	1.18	0.24				
2 seconds	Simple	-0.15	0.88	0.18	.75	46	.46	.02
	Complex	0.03	0.78	0.16				
5 seconds	Simple	-0.17	0.88	0.18	.66	46	.51	.02
	Complex	-0.01	0.78	0.16				
8 seconds	Simple	-0.11	1.08	0.22	.51	46	.61	.01
	Complex	-0.02	1.00	0.20				
15+ seconds	Simple	-0.10	0.81	0.16	.33	46	.74	.005
	Complex	-0.02	0.88	0.18				

*Significant effect

The results of the independent samples t-tests indicated that none of the complexity pairs of mean of means z-scores on the eight timing intervals differed from each other, although the simultaneous condition, as might be expected, approached significance. It is noted that the independent samples t-test was used for the analysis as a more conservative test since half of the data for each mean z-score came from study 1 and half from study 2. Figure 20 shows the relationships among these figures of merit. Recall that lower values mean better performance.

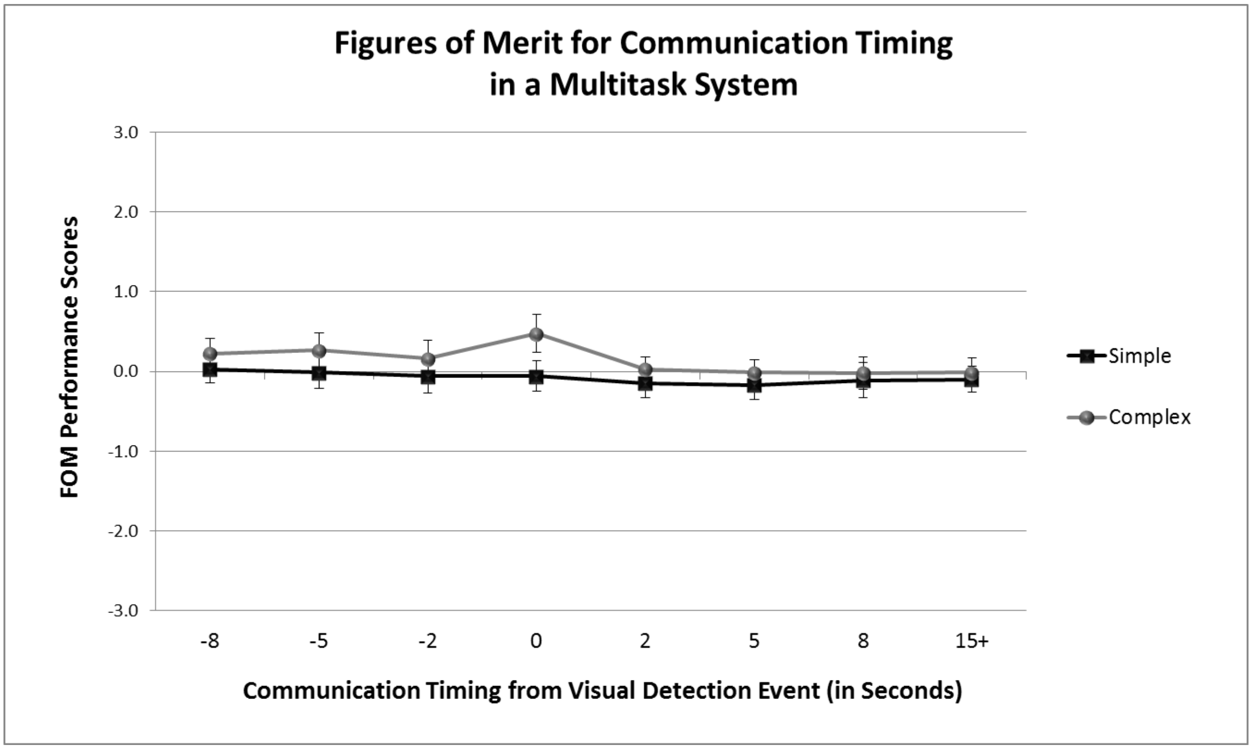

Figure 20. Figures of merit for timing intervals by communication complexity.

Finally, the mean of these derived z-scores and standard deviations were used as data for an overall figure of merit analysis to evaluate the impact of simple versus complex communications in this multitask system. Since both a measure of central tendency and a measure of variability in figure of merit analyses had been found in a prior study to yield different conclusions about the quality of overall task performance (LeMay & Comstock, 1990), the same evaluation approach was used in the current research.

The figure of merit data for overall task performance can be found in Table 20. (Recall that lower z-scores for means equate to better performance). Two independent samples t-tests compared the figures of merit for the combined z-score means and standard deviations for task performance for the two levels of communication complexity in the system. The t-test comparing the two means was significant, t (14) = 3.21, p = .006 (two-tailed), η^2= .60, a large effect. When all of the performance data is combined as a figure of merit to define task performance in a multitask system, complex communications (M = 0.14, SD = 0.97) led operators to perform with significantly less overall efficiency in the system compared to simple communications (M = -0.08, SD = 0.93). The t-test comparing the two standard deviations was not significant (p =.52). No differences were found when comparing the overall variability in task performance in the presence of simple (SD = 0.93, sd = 0.09) versus complex (SD = 0.97, sd = 0.15) communications.

Table 20

Figures of merit data and t-tests comparing simple and complex communications

Communication Complexity	*M*	*SD*	*(sd)*	*t-test*	*t*	*df*	*sig. (2-tailed)*	*η2*
Simple	-0.08	0.93	0.09	Mean	3.21	14	.006 *	.60
Complex	0.14	0.97	0.15	SD	.65	14	.520	.06

*Significant effect

This figure of merit analysis demonstrates an alternative method for analyzing task performance data in a multitask system. Based on this analysis, it appears that the largest contributor to differences in task performance was the complexity of communications. Using this method, the timing of communications had relatively minor impacts on overall task performance. Important trends that might prove worthy of further study became apparent with the figure of merit approach, but otherwise was hidden using traditional statistical methods.

A summary of all research results is shown in Table 21.

Table 21

Summary of research results

	Measures	Results
STUDY 1: **COMMUNICATIONS BEFORE MALFUNCTIONS**		
Simple vs. Complex Communications	Response time	Complex ≈ 2/3 sec longer
	Detection accuracy	Complex ≈ 80%; Simple ≈ 88%
	Tracking	NS; small effect
	Workload	Complex ≈ 6 points higher
Timing x Complexity	Response time	0 sec. complex ≈ 1.3 - 1.6 sec. delay
		2 sec. complex ≈ 1 - 1.15 sec. delay
		No sig. differences among simple conditions
	Detection accuracy	0 sec. complex ≈ 20% degraded
		No sig. differences among simple conditions
	Tracking	NS; small effect
	Workload	All complex ≈ 6 points higher
Limits of echoic memory	Response time	8 sec. simple vs. complex NS; small effect
	Detection accuracy	8 sec. simple vs. complex NS; small effect
	Tracking	NS; small effect
	Workload	8 sec. complex ≈ 8 pts > 8 sec. simple
STUDY 2: **COMMUNICATIONS AFTER MALFUNCTIONS**		
(Tests alternative assumptions: Context,	Response time	NS complexity or timing ; small effect
anticipation of communications, auditory	Detection Accuracy	NS complexity or timing ; small effect
fatigue, mere presence of communications)	Tracking	NS complexity or timing ; small effect
	Workload	2 sec. complex > all simple;
ADDITIONAL FINDINGS		
Individual differences	Tracking	Females' tracking more variable than males'
Figure of merit analysis	Trends across studies	Auditory distractions worse -2 to +2 sec. near visual event
		Complex communications most disrupting
	Overall estimate of efficiency of system	Communication complexity most costly to operator performance; timing less important

CHAPTER FIVE: GENERAL DISCUSSION

The current research evaluated the impacts of complexity and timing of communication tasks on operator performance of a visual detection task and subjective workload in a multitasking environment. While many studies have examined operator performance in dual-task scenarios, fewer studies have examined these variables under multitasking. Laboratory research using such paradigms is needed in the current era of the rise of multitask-laden technologies.

While the term "multitasking" is used in popular culture, human multitasking has been termed a myth (Loukopoulos, et al, 2009) because operators typically perform task switching, or alternating between an ongoing and an interrupting task. Task switching takes various forms such as interleaving an ongoing and an interrupting task, or fully completing one task before another. The context of the operations is a defining factor. In task interruption research, the resumption lag—or the time it takes an operator to resume an ongoing task after an interruption, has received much attention (Trafton & Monk, 2007). Less is known about an earlier point in the theorized task interruption timeline, the interruption lag—or the characteristics and influences the initial interruption has on the rest of the timeline.

In addition, auditory tasks have not been well studied in the task interruption research domain. It is known that an auditory stimulus has a life of about 2 to 5 seconds (Treisman, 1964) and after this time, the stimuli must be processed to completion, mentally rehearsed to retain it in working memory, or allowed to decay. Visual memory has a span of up to one second (Sperling, 1960), no doubt owing biologically to the relative length of most visual events.

One information processing theory that integrates auditory and visual modalities is Multiple Resource Theory (MRT; Wickens, 2008). A principle within MRT is that an ongoing visual task may effectively time-share with an interrupting auditory task with little performance cost to either. Research has supported that assumption (e.g. Wickens, Goh, et al, 2003).

However, research in multitask contexts has found situations where an auditory task captures attention at the expense of performance of an ongoing visual task, a term referred to as auditory preemption. One proposed reason is that that the onset of the auditory stimulus commands the operator's full attention thus disrupting performance in the visual task. Another is that the operator strategically choses to focus on an auditory task of some complexity so as not to lose the message. Underlying the second reason is that the operator understands their limitations and spreads attention strategically across a finite pool of cognitive resources. Research has supported both of these ideas (e.g. Latorella, 1998). Mixed results have been found regarding the situations under which MRT or auditory preemption applies (Lu, et al., 2013).

This gap in the literature led to the design of the current research. Few studies have manipulated both the timing (in terms of the limitations of the echoic memory store) and the complexity (in terms of an onset vs. strategic demands) of an auditory interruption in a multitask context within one study. In addition, few have offered the opportunity to examine alternative factors that may explain auditory attention capture principles.

Study 1 examined the impact of complexity and timing of communications occurring before or simultaneous to critical ongoing visual detection events. Communication timings before visual detection events were chosen in order to examine the onset (defined as a simple communication) versus strategic (defined as a complex communication) auditory preemption dichotomy as a function of the limitations of echoic memory.

It was expected that communications occurring simultaneous to or within 5 seconds before a visual detection task would be most disruptive to the operator (confirming onset preemption). Complex communications were expected to result in worse performance outcomes and higher subjective workload compared to simple communications, supporting a strategic preemption principle. Beyond 5 seconds, similar impacts were not anticipated, supporting MRT principles.

The communications in this study were purposely kept short in order to examine the two auditory preemption assumptions, as well as considering prior task interruption research that interruptions longer than 13 seconds exponentially increase resumption lag for the ongoing task.

In Study 2, communications were placed at similar timings as Study 1, but after the visual detection task, thereby offering the opportunity for alternative assumptions that the context of the task may be the factor in any visual task performance declines. For example, in situations where communication interruptions are expected, there may be a cost to task performance while the operator waits in anticipation of the upcoming auditory event. Such anticipation may result in a type of auditory fatigue that has a cumulative attention capturing effect. Also since various types of auditory interruptions have been found to impact transportation tasks, the mere presence of auditory communications may influence visual task performance. Under these assumptions, Study 2 results were expected to reflect those from Study 1.

Communications Occurring Before System Malfunctions

As shown in the results summary in Table 21, complex communications occurring before a visual detection event more significantly impacted response time to and accurate detection of the malfunction compared to simple communications. Overall, complex communications resulted in response time delay of about 0.67 seconds and exacted about an 8% cost in terms of detection accuracy compared to simple communications. Workload also was rated about five to eight points higher, overall, for complex communications compared to simple ones.

When the impact of the proximity of the communication to a visual detection event is considered, the most severe performance impacts are seen when a complex communication occurs simultaneous to or within about two seconds before or after the malfunction to be detected. In the worst case, complex communications occurring within two seconds or less of a visual detection event required fully 1.3 to 1.6 seconds longer for response compared to when simple communications occurred at two seconds or longer before the malfunctions. This delay in seconds mirrors prior research regarding a delay of pedestrians in crossing a street when talking on a cell phone (Neider, et al., 2010).

The complex simultaneous condition also exacted the largest accuracy cost of any of the other conditions, which did not differ from each other. When communications and visual detection events occurred together, operators were more than 20% less accurate in detecting malfunctions compared to the condition where simple communications occurred five seconds or more prior to the malfunction. This finding for a decrease in task accuracy in the presence of auditory interruptions supports prior research (e.g., Lu, Wickens, et al., 2013). Additionally, operators appeared aware of the workload impact of complex communications over simple ones. However, there was no evidence from this study that they could discriminate among the timing of the complex communication in relation to the malfunction.

Unfortunately in study 1, the tracking task results did not indicate an impact of communication complexity or timing. Those null results may have been due partially to a data analysis or software design concerns, which are discussed further in the limitations section.

Communications Occurring After System Malfunctions

Also as shown in the summary table, competing assumptions regarding the findings from this research were not supported. For example, it did not appear that the expectation of an upcoming communication or the mere presence of communications along with the study tasks would be responsible for the results from Study 1. In addition, an argument for a cumulative effect on performance of study tasks due to complexity or timing of communications, such as from auditory fatigue across the experiment, did not appear to be an explanation for results seen here. If those factors had been in effect, the findings from Study 2 would have been expected to reflect those in Study 1.

However, the workload findings from Study 2, and also compared to Study 1, are intriguing. The findings from the two studies both indicated that participants rate their workload as higher when complex communications are involved but they are little able to discriminate the conditions in which their performance might be the most severely impacted. Participants rated their workload as higher, overall, in Study 2 compared to Study 1. While a larger variability among the workload scores in Study 2 may account for some of this difference, it does not account for all of it. The workload ratings of participants were higher in every comparable condition of Study 2 compared to Study 1. In fact the highest rated workload condition in Study 2—a communication event occurring two seconds *after* a visual detection event—was rated about eight points higher than the highest rated condition in Study 1—the condition where the communication and visual detection event occurred simultaneously. These findings are discussed in more detail later in this discussion.

Additional Findings

The additional analyses for sex differences found an effect for the tracking task such that males generally performed better than females and with less variability overall, at least for certain of the complex communication conditions. In addition, the pattern of scores for females suggested that their performance for a visual-spatial task in a multitask system appeared to somewhat improve with the additional task loading of communications. However, this does not make intuitive sense, and further research would be needed to investigate this effect.

Several additional analyses were performed using a figure of merit (FOM) approach. FOMs are often used in engineering design or business applications to develop a criterion that may be used to estimate the efficiency or usefulness of a multitask system. This approach provided two useful analyses: Observations of trends across the data from study 1 and study 2, and the development of the overall figures of merit to assess both the timing and the complexity of communication interruptions on visual detection tasks.

Of particular note is that the trends analysis found an obvious perturbation in the performance of system tasks in which the trend lines diverged between the -2 seconds and +2 seconds intervals. While the trend could be detected in the statistical hypothesis testing for the study measures individually, the plots of these lines helped to visualize these effects across and between the measures. In addition, these graphs showed the wider variability of task performance in the presence of complex communications compared to simple communications.

Despite these observed trends, the overall FOM's computed for each of the four timing intervals at each of the two levels of complexity indicated an overall relatively small effect of the impact of timing of communications on task performance. However, the FOM derived as a measure of communication complexity found a large effect for the impact of complex communications on the performance of system tasks compared to simple communications. Using this approach, it may be stated that it is the complexity of the communication, not necessarily the timing that has the largest potential to disrupt task performance, even for the very controlled, closed-loop type communications that were the basis of this study. The FOM approach was highly useful in this research and is recommended in future evaluations of multitask systems.

Theoretical Implications

Auditory Preemption, Multiple Resource Theory, and Echoic Memory

This research expands existing knowledge regarding some of the characteristics that set onset preemptions apart from strategic preemptions (Wickens & Liu, 1988; Latorella, 1998). When a communication is brief (less than about 10 seconds) and the required response is also brief (less than about 15 seconds) there is little impact to interruptions of ongoing visual tasks from simple communications. Such brief responses might include activities such as repeating back a request without a requirement to remember it (i.e. reading the time, reporting an altitude), or a giving a response that may be contained in long-term memory (i.e. stating your name or an aircraft identification, reporting the last observed event). This would indicate a null effect for onset preemption under these very specific situations. In a very brief, simple communication that requires a well-defined response, the attention-capturing effect appears minimal, if it exists at all.

There was also little interruption when brief, simple, well-defined communications occurred at about 8 seconds or more from a visual task. This finding supports the Multiple Resource Theory (Wickens, 2002) assumption that certain visual and auditory tasks may productively time-share cognitive resources. The boundaries of this assumption may well exist up to the limits of the auditory sensory store plus any associated information processing requirements for a particular message.

Communications such as those described above are common in the controlled aviation communications environment. The finding may also have implications for other environments such as surgical theaters in medicine, emergency response situations, or other situations in which communications are well-known or easily learned.

Strategic preemption appeared involved for most types of complex communications (i.e. involving information processing requirements), and especially when the communication occurred very close to an event requiring visual detection, defined here as simultaneous to or within five seconds of a visual detection event. This finding applies even when the communication is brief (less than about 10 seconds) and the required response is also brief (less than about 15 seconds). This finding reflects the research regarding inattentional blindness (e.g. Neisser & Becklen, 1975) that has also been supported in transportation research (e.g. Strayer, et al., 2003).

Considering these findings for onset versus strategic preemptions overall, the results must be considered mixed, and that has been an ongoing theme in studies attempting to clarify these two constructs. This research appeared to support a finding for strategic preemption when complex communications (those which impose information processing demands) are involved. This was especially the case when the complex communication occurred simultaneous to or within five seconds of a visual detection event. In those cases an operator must make a decision to either attend to the communication or complete the visual detection task, creating a conflict.

The communications task data in this research indicated a near ceiling effect with most participants completing the communication responses correctly. This indicates that the cognitive resource-sharing cost came at the expense of the visual task as participants made decisions to complete the communication, perhaps not being aware of the cost to other task performance. Perhaps most convincing is that if onset preemption had been at work, there should have been some observable effects or trends on task performance among simple communication conditions. For example, since onset preemption is defined as the attention-capturing quality of an auditory event at the expense of performance of visual or manual tasks, then the most pronounced effects on performance should have been in the response times and accuracy data of the simultaneous or 2 second simple conditions compared to the 5 and 8 second simple conditions. No such effects were observed in the performance data for simple communications.

Other Theories of Memory

The current research was designed based on information regarding the capacity of the auditory sensory memory stores and theories of attention. However, theories regarding other aspects of memory may also explain these results.

For example, the distinction between prospective versus retrospective memory may be important to consider. Retrospective memory is the oldest and most well studied of the two, with various subdivisions proposed such as episodic versus semantic memory (see Tulving, 2002 for a review) and implicit versus explicit (Graf & Schacter, 1985). However, all have in common the aspect of retrieving information, such as by free recall or recognition of past events, objects, words, etc. A common retrospective memory task is word list recall.

On the other hand, prospective memory is a newer concept, distinct subdivisions are not well-defined, and there may be aspects that are not dissociable from retrospective memory for some tasks (Graf & Uttl, 2001). However, one distinction regarding types of prospective memory includes event-based or environmentally-cued tasks (such as pressing a key when a target event occurs), and time-based or self-cued future actions (such as remembering to press a key in 10 minutes; Smith, Della Sala, Logie, & Maylor, 2000, p. 312. See also Einstein & McDaniel, 1990). In general, prospective memory involves remembering to perform a task in the future. Prospective memory tasks may include monitoring a tea kettle, remembering to convey a message to someone, or remembering to take nightly medications (Graff & Uttl, 2001, p 440).

How do these memory concepts relate to the current research? In terms of the communication task, the task demands were new only for two blocks (one for simple and one for complex communications). After that, communications in remaining timing intervals essentially became retrospective memory tasks, at least for the memory of the process of the tasks. (None of the words were repeated so this was not a traditional word list recall task). It might be expected that performance would be worse for those two initial timing conditions (i.e. simple and complex) for each participant compared to the remaining three for each complexity level. The data was not analyzed in this fashion but would be an interesting additional analysis.

In addition, in the complex communication tasks (where the participant had to generate three words starting with the last letter of a word prompt), each participant was given 10 trials for each timing interval, or a total of 40 word prompts across the experiment. The study design did not control for the number of times the last letter was the same. So for example, the letter "p" occurred as the last letter in a word prompt for two of the words, whereas the letter "s" occurred as the last letter for five of the words. Several participants were noted to use the strategy of trying to recall words they had used before, making those subsequent trials a type of retrospective memory task. Further analysis of the response times and accuracy in responding to visual detection tasks for the trials which used the more frequently occurring letter prompts may also help to further describe the results here.

Apart from the communications task, the system monitoring task is arguably a prospective memory task with environmental cueing (i.e. participants were cued by the gauge offsets to press a key to reset the gauge). So an interpretation of the current study results is that an event-based, short-term, visual prospective memory task is more disrupted when a retrospective auditory memory task is implicit (i.e. in the complex communications, or freely recalling any word that comes to mind) compared with an explicit auditory task (i.e. in the simple communications, or immediate recall of the auditory cues). Said another way, event-based, visual monitoring, prospective memory tasks are more degraded under conditions of auditory free recall rather than repetition.

A theory of working memory (Baddeley, 1986; 2003) may also assist in explaining these results. Baddeley's integrative model of working memory proposes three primary components: A visual-spatial sketchpad, a phonological loop, and a central executive which acts a control system to gate and manage incoming information.

Baddeley (2003) explained that the phonological loop evolved to facilitate language acquisition. The phonological loop may be considered a short-term memory processing center for auditory information. Once an auditory input enters, it is acted upon through sub-vocal rehearsal to retain it in the loop, and this rehearsal holds the information for further processing, such as producing a spoken output or interfacing with long-term memory. According to a summary by Sternberg (2006, p. 170), without this capacity for sub-vocal rehearsal and short-term retention, auditory information decays in about 2 seconds. This capacity is several seconds shorter than the original studies by Treisman (1964) that placed an outer limit at about 5 seconds.

The results of the current study appear to support this 2-second limit for the articulatory loop. As was seen most clearly in the figure of merit, obvious perturbations for performance measures occurred between the -2 to +2 seconds timing intervals when a communication was paired with a visual detection task. In the current research, it appears that the interruptions imposed by a visual detection task may have acted to interrupt the articulatory loop, at least for complex communications.

Baddely's (2003) model as proposed is essentially one of working memory. It interfaces with long-term memory as auditory information, words in particular, become longer or when word meanings become important. Further relating Baddeley's model to the present research, it is likely that the simple communication conditions made use only of the articulatory loop as the individual simply attended to the words sub-vocally and repeated them from the working memory store. However the complex communications required the individual to first understand and process the request, using the articulatory loop, and then switch to long-term memory to generate a word that matched the demands of the request. Therefore the current research could be said to be a comparison of auditory working memory (i.e. simple communication requests) and long-term verbal memory retrieval (i.e. complex communication requests) processes when the secondary interrupting task is a visual detection event.

Subjective Workload

Several participants commented after the study that they were surprised with the workload of the complex conditions. While they were able to detect an increase in workload with complex communications, they were not able to accurately determine which timing conditions contributed to workload. In the condition where visual task response time and accuracy did not differ when simple and complex communications were present, participants nonetheless rated the complex condition higher. This finding for a dissociation between subjective workload and actual task performance mirrors prior findings of individuals' poor post-task estimates of their performance (Vincenzi & Mouloua, 1998; Lesch & Hancock, 2004; Strayer, et al., 2003), and lower reliability for anticipating the cognitive demands of a task compared to physical demands (Sublette, et al., 2009).

In the post-experiment phase, participants reported that during the complex conditions, they had reflected about their past responses to complex communication requests which added an extra (and unanticipated) cognitive workload to the task. This observation indicated that a metacognitive process had been activated in the context of the complex communications task and the subjective experience of workload may be the best indication of this observation.

What was unclear from these post-experiment observations is whether the task demands activated the metacognitive process, or whether decisions to strategically delay performance of the visual task while dealing with the auditory task activated metacognition. Recent research with air traffic controllers indicated that the relationship between mental workload and task demand is dependent on controllers' capacity to understand and manage their own metacognitive processes (Loft, Sanderson, Neal, Mooij, 2007). In addition, other research has found evidence for "supertaskers" whose driving performance does not decline as expected when coupled with an auditory memory span task (Watson & Strayer, 2010). Future research would help to further define these interrelationships among metacognitive processes and workload in the presence of communications while multitasking.

Practical Implications

As future complex technological systems become increasingly automated, it will be important for designers to have information regarding when and how both auditory and visual tasks interrupt each other and how automation may assist in relieving some of the bottlenecks and interruptions. The results from this research should contribute detailed, micro-level data toward a better understanding of design principles when communication tasks are involved.

In terms of aviation communications, such detailed data regarding when a communication interrupts critical tasks on the flight deck may be important in several ways. For example, it may provide recommendations for training of pilots and air traffic controllers regarding how to cope with communications that preempt other tasks and how to manage other flight tasks when responding to communications. It may also provide details for the design of future DataLink systems that are planned for the NextGen air transport initiative.

Such detailed information regarding the impact of communication interruptions to other, ongoing tasks also has broader implications beyond aviation. For example, the information may be useful for designing voice directions for GPS systems used in aircraft or automobiles, such as when they should interrupt the operator based on the complexity and timing of other critical tasks that must be performed within the system. Detailed information regarding communication interruptions may also be useful in the design of nuclear power plant, railway, industrial, or emergency management communication systems.

The information found in this research regarding the complexity and timing of communication interruptions may also assist in areas of training for team communication and coordination, and in communication in medical environments such as in surgery or in the emergency department. For example, there appears to be a critical window of time in certain visual tasks where the primary operator should not be interrupted with questions or auditory requests. Training co-operators to be alert to this critical time window may help to reduce interruptions and subsequent errors at critical times in a visual task.

Therefore, based on this research, the following practical suggestions are offered.

1. In operational situations involving experts, or when communications involve well-known or well-defined topics, keep communications requests brief. Less than 10 seconds for any request or required response is best.

2. When possible, break complex communications into simple, well-defined ones during task requiring visual monitoring.
3. Space communications apart to allow for a decrease in task load during visual detection task performance.
4. Anticipate upcoming visual detection task loading and defer complex communications until after the task has passed.
5. When possible, anticipate extra time for responding to visual detection tasks when communications are occurring and plan for unintended consequences of the lost time. For example, increase temporal distance (at least 2 seconds) between tasks to be completed or between moving objects.
6. Simply asking the operator about their amount of workload is not enough to determine the impact of communications on performance of visual detection tasks. Spot checks of important performance outputs are required.
7. Expect as much as a 20% decline in visual detection accuracy during co-occurring communication tasks and plan for checks of performance when tasks are critical.
8. In systems design, allow at least 10 seconds to pass after the operator responds to a communication before delivering a visual signal that requires detection.

Limitations of the Research

In this research the response time measures were the most sensitive to study manipulations, followed by visual detection accuracy. The tracking task data was not sensitive to the independent variable manipulations in either study 1 or study 2. Since prior research, especially from the driving domain, has indicated an effect on tracking tasks (i.e. lane-keeping), the insensitivity of the tracking data from this study presented a limitation on interpretation and comparison of study results to other studies in the transportation domain. There are three primary factors that may contribute to this null finding.

First, it is possible that participants always maintained accuracy for this task regardless of communication or the system monitoring events that occurred. Since this was the only fully manual task, the "hands-on" nature of the task may have commanded participants' full attention.

There may also be a software explanation. Several participants reported a lag in the inputs of the tracking task as well as a relative ease in performing this task. That is, the task was not highly sensitive to inputs and often participants could take hands off the task for several seconds before inputs were required to maintain the target within the specified box.

Finally, a statistical explanation is possible. The values for the root mean square error from center for the tracking task data was averaged over the entire timing block rather than only near the communication presentations. This averaging method may have introduced noise in the data that may have obscured any differences. Further analyses of the tracking data may help to determine which of these factors may have influenced these results.

An additional analysis of the tracking task data for sex differences did find some modest effects for the complex conditions. It was noted overall that females performed the tracking task generally with less precision and more variability than males. Future studies of tracking task performance may help to further define these spatial/temporal differences using sex as a variable.

Another limitation involves the communications task. This task was based on the theoretical construct of degrees of difficulty of processing verbal information in the auditory modality. It did allow for precise manipulation of the information processing construct as applied to comprehension and processing of a verbal requests and oral output from the request. This resulted in a tighter definition and control of the "simple" versus "complex" communication manipulation. However, these were not “real world” communications. While information processing tasks have been used frequently to represent communications in basic research in the transportation domain, they do present limitations to statements regarding generalizability to situations were more complex or free-flowing conversations occur.

Therefore, communications evaluated here were relatively predictable and brief, similar to closed-loop communications used by experts in several domains. These findings may not apply fully to communications that are more complex, unpredictable, or that require several rounds for clarification of the message. Such communications could be expected in emergency situations were an understanding of the situation and context are required. These brief and predictable communications may account for the lack of findings for onset preemptions. It is possible that the onset preemption phenomenon exists for communications that are unexpected and does not exist in context where a communication and the required response can be relatively well anticipated. Nonetheless, the findings here represent the “lower boundary” of the impact of auditory communications to ongoing visual tasks, and the impacts were found to be significant.

Also this study used a young, relatively computer-savvy group of participants. Either of those factors may work for or against the results found here. For example, a younger sample may actually produce worse performance results than an older sample due to the relative naiveté of the participants to the goals of the research. That is, an older group may have been able to “guess” the goals of the study thus positively influencing results. Alternatively, the younger groups’ relative comfort with computers may have resulted in better outcomes which may be reflected in these study results. Arguing in the other direction, older adults are known to have slower response times to most stimuli, a function that has been shown to be exponential past about the age of 50. It is possible that an older adult subject pool would have exhibited worse results on study tasks.

Future Research

Several suggestions for future research have been generated from this study. For example, a study with a similar design as the current one, but using more realistic communications, would assist in determining the generalizability of information processing communications tasks from the more time and cognitive resources-intensive real world conversations. And as stated in the limitations, it would prove interesting to complete the current study with an older adult sample to provide a comparison of the impacts of communications on performance of visual detection tasks across the life span.

With regard to the functional capacity of echoic memory, the current research results suggest that beyond about five seconds, the operator may be able to resist the impact of communications on visual detection tasks under similar communication circumstances as those presented here. This would support multiple resource theory (Wickens, 2002) principles regarding resource-sharing of auditory and visual tasks, at least for these relatively straight-forward communications. It appears that with less than the five seconds between a discrete communication and a visual detection task, assumptions of auditory preemption theory (onset and strategic preemptions; e.g. Latorella, 1999) may apply. Several studies to vary levels of real-world, simple versus complex communications considering echoic memory limitations would help to further define these relationships.

The workload results from this study, combined with informal participant observations after the study about the surprising difficulty of the communications task, indicate that something about the context of the complex communications task activated metacognitive processes.

However, it is unknown which came first: The retrospective memory of prior performance in the earlier complex conditions which caused participants to reflect on their ongoing cognitive processes, or their decisions to strategically delay visual task performance at the service of completing the communication tasks. Future research to examine the metacognitive processes involved in auditory preemption versus multiple resource theory in interruption management (e.g., Wickens, Dixon, & Seppelt, 2005) using a multitasking paradigm is recommended to further explore these observations.

Prior results from the transportation domain have found impacts on similar navigation tasks, such as lane-keeping, from communications (e.g. Horrey and Wickens (2006), and the results from study 1 indicated a medium effect, though no statistical significance. As was demonstrated in the a posteriori analyses, the addition of sex as a covariate improved the statistical model and indicated that this individual difference is an important variable to consider in future research regarding tracking task performance. In addition, the analysis found a significant interaction of sex differences and timing, which also supports recent calls (i.e. Hancock, 2011) to consider the perception of time by sex as an important variable in research studies.

The figure of merit approach provided depth to the data analysis and understanding of interrelationships among the independent and dependent variables in this research. It allowed for a different visualization and a rich understanding of the data when combined with the traditional statistical hypothesis testing analyses. This approach is recommended in future evaluations of multitasking systems.

APPENDIX A:
UCF APPROVAL OF HUMAN RESEARCH

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Sally A. Stader**

Date: **March 18, 2014**

Dear Researcher:

On 3/18/2014, the IRB approved the following human participant research until 3/17/2015 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Conversations in multitask environments
Investigator:	Sally A Stader
IRB Number:	SBE-14-10074
Funding Agency:	
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 3/17/2015, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Joanne Muratori

IRB Coordinator

APPENDIX B:
INFORMED CONSENT

Conversations in Multitask Environments

Informed Consent

Investigator: Sally Stader, UCF doctoral student in human factors psychology

Faculty Supervisor: Mustapha Mouloua, Ph.D., professor in psychology

Investigational Site: UCF Psychology Building, Room 307 and Suite 303.

Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being invited to take part in a research study which will include up to 100 people. To continue, you agree that the following applies to you:

- You are 18 years of age or older.
- You have 20/20 vision or wear glasses or contact lenses so that your vision is near 20/20.
- You have normal color vision.
- You can operate a standard keyboard and joystick.

What you should know about a research study:

- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: This is an experiment that examines conversations during multitasking. During the study, you will perform tasks on a computer and provide answers to conversations. After several minutes on the tasks, you will be asked to provide ratings of various aspects of mental workload that you experienced during the tasks.

What you will be asked to do in the study: The total time for this study is 2 hours. Here is what you will do during that time:

1. Receive an explanation of the study and complete this consent process (10 min.)
2. Receive training on experiment tasks (20 min.)
3. Complete several short sessions engaged in study tasks and conversations followed by a short survey of your workload after each session (about 1 hour, 15 minutes)
4. Complete a post-study demographic survey and exit (15 minutes)

Location: This study takes place in the Transportation Research Group lab in the UCF Psychology Building, Room 307 or in Suite 303.

Time required: Full completion of this study requires about 2 hours.

Risks:

There are no reasonably foreseeable risks or discomforts involved in taking part in this study.

Benefits:

As a research participant you will not benefit directly from this research, besides learning more about how research is conducted.

Compensation or payment:

Participants may expect to spend up to 2 hours performing experimental tasks for which they will receive course credit for the amount of time they participate. Course credits at the rate of 1.0 credit per hour (.5 credits per ½ hour) are recorded in the psychology department Sona system. There is no cash payment offered for this study.

Confidentiality:

Your personal, demographic data collected in this study is limited to people who have a need to review this information. Your identity will be kept confidential. Your name is not collected. Information you provide will be assigned a code. All of the information from the study will be kept in a locked filling cabinet or stored on a password protected computer. Your information will be combined with information from other people who took part in this study. When the researchers write about this study to share what was learned with other researchers, they will write about this combined information. Your name will not be used in any report.

Study contact for questions about the study or to report a problem:

If you have questions, concerns, or complaints, or think the research has hurt you, talk to Sally Stader, psychology doctoral student, (863) 712-2497 or by email at sally.stader@knights.ucf.edu or Dr. Mustapha Mouloua, Faculty Supervisor, UCF Psychology Department at (407) 823-2091 or by email at mustapha.mouloua@ucf.edu

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

UCF University of Central Florida IRB
IRB NUMBER: SBE-14-10074
IRB APPROVAL DATE: 3/18/2014
IRB EXPIRATION DATE: 3/17/2015

APPENDIX C:
EXPERIMENTER SCRIPT

Communications in Multitask Environments—Experimenter Script

<u>MATB-II</u>

Prior to Subject Arrival

I) On each computer, check that joystick, keyboard, mouse, and headsets are connected
II) Put new covers on the headsets
III) Turn on noise cancelling switch on headphones
IV) Turn on the MATB computers
V) Log in to one of the COS computers with your log in or with the log in posted at the front computer
VI) Log in to Sona; Find participant's Sona number and write it on the Run Sheet & Participant Sheet.
VII) Fill in date, time, and system to be used on the Run Sheet for the participant
VIII) Complete the Participant's Sheet with above data
IX) Place the following next to each computer:
- A) Consent form
- B) Demographics form
- C) NASA TLX scales description sheet
- D) Pencil to complete Demographics

X) Take out a Study Feedback form and a Psych Participant Evaluation form for end of experiment
XI) Prepare the MATB-II experiment computer
- A) Check that the NUM LOCK key is on
- B) Click on the MATB-II icon on the desktop
- C) In the "MATB: Please Select the Events File" dialogue box, select "**MATB_Events-TRAIN**" file
- D) Click "Open".
- E) Click "File" at the top, then click "Start"
- F) Press keys 1 through 8 to activate the Flow Rates values on the bottom right of the screen.
- G) Move pointer to bottom of the screen to bring up the task bar
- H) Click on the task bar to pause the MATB-II tasks
- I) If this does not work, click on the far edge of the screen and outside the task area to pause
- J) Hold until ready for Training session

Subject Arrival and Introduction

I) Welcome participant- Read verbatim:
- Good (morning/ afternoon). Thank-you for participating today.
- This is an experiment that examines communications during multitasking.
- During the study, you will perform tasks on a computer and provide answers to communication requests.
- After several minutes on the tasks, you will be asked to provide ratings of various aspects of workload that you experienced during the tasks.
- After the training portion, the experiment today takes about one hour.

II) Informed Consent
- Please read this consent form for UCF's Institutional Review Board (IRB). This form outlines your rights as a research participant. *(Allow a few minutes)*
- Do you have any questions?
- Please turn your cell phone off at this time.

III) Study Summary
- Here is what you will be asked to do in the study.
    - First, you will receive a short training on tasks using the automation software.

- You will operate the system using a standard keyboard, mouse, and a joystick.
- You will hear the communications over headsets. The covers on the headsets have been provided just for you.
- After training, you will fill out a brief the workload rating scale.
- Then you'll complete a short demographic form.
- This training takes about 20 minutes.
- We will have time for a break if you need one after the training.
- After the break, the experiment takes about 1 hour.
- Because we are piloting this study, you will receive 2 credits for today's experiment.

- Do you have any questions?

Training Session with NASA TLX and Demographics

I) Explaining overall task goals and equipment

- **General introduction:** For this study, you will be performing three tasks using the computer system in front of you. They are called "system monitoring", "tracking", and "resource management" *(point to or explain the area on the screen).*

- **Training—Resource management:**
  1) The resource management task represents a fuel management task.
  2) These large rectangles, labeled A, B, C and so on, are the fuel tanks. The green shows the tank levels.
  3) The numbers near the tanks are the amount of fuel in the tanks.
  4) The small rectangles, labeled 1, 2, 3, and so on are the pumps. They transfer fuel between the tanks. The arrows show direction of fuel flow.
  5) Take a moment to locate each of the tanks and pumps.
  6) The column to the right shows the pump flow rates per minute.
  7) Some people find it useful to know those flow rates. You can decide if they help you.
  8) <u>The main goal of this task is keep the fuel tanks at their optimum levels by transferring fuel between the tanks.</u>
  9) You will use the number 1 through 8 keys to turn the fuel pumps on or off.
  10) You may use either the row of number keys or the number pad for these keys
  11) The pumps are green now which means they are on and moving fuel to the tanks.

  12) **Here's how to perform the Resource Management task.**
      a) Tanks A and B are the main tanks. They should stay in the optimal range where you see them now. The optimal range is between 2000 and 3000 units.
      b) The tanks are now at about 2500 units.
      c) The small grey knobs on each of the tanks show the optimal range.
      d) Tanks A and B should not be allowed to be completely empty or full at any time. If they do run empty or full, their pumps will shut off.
      e) Tanks C & E supply fuel to Tank A.
      f) Pumps 1 & 2 take fuel from Tanks C & E to Tank A.
      g) Pump 5 takes fuel to Tank C from Tank E.
      h) Take a moment to study the fuel flow for Tank A.
      i) The set-up is the same for Tank B.
      j) Tanks D & F supply fuel to Tanks B.
      k) Pumps 3 & 4 take fuel from Tanks D & F to Tank B.
      l) Pump 6 takes fuel to Tank D from Tank F.
      m) The pumps 7 & 8 transfer fuel between Tanks A & B.
      n) Finally, it is important to know that Tanks C & D also should not be allowed to be

completely empty or full. If they do run empty or full, their pumps will shut off.

o) Tanks E & F never run out.
p) Again, to do this task, click keys 1 through 8 to transfer fuel between the tanks.

13) Do you have questions about this task?
14) Let's try it. *(Click inside the screen on the laptop if the run is activated. Otherwise, click "File" at the top, and "Start", then "Yes" button or "Enter" key.)*
15) *Monitor the 2 minutes of the RESMON task. Correct errors. Instructing about basic pump and tank operation is ok. But, don't give a strategy for managing fuel*
16) *At 2:00 mark, the RED light will come on.*
17) *Move cursor to bottom and click on task bar to pause MATB-II—or click outside tasks area*
18) So summarize, the main goal of this task is to keep A and B in the range between 2000 to 3000 units
19) Use tanks C, D, E, and F to keep the main tanks A and B in the optimum range.
20) Don't let any of the tanks A, B, C, or D get too full or empty.
21) Tanks E and F never run out

- **Training—Tracking:**
  1) The tracking task represents a navigation task.
  2) Are you right-handed? *(Place joystick on side of handedness).*
  3) You will use this joystick to keep the round tracking symbol in the middle of the target box
  4) The symbol won't stay in the center on its own. Move the joystick up or down, or side to side, to move the symbol to the center.
  5) <u>The main goal of this task is to keep the symbol in the center area.</u>
  6) Try it. *(Restart the task by clicking inside the task area on the screen.)*
  7) *After about 30 seconds on Tracking, ask the participant the following:* Now look at the Resource Management task to see if anything is needed.
  8) *Instruct participant on RESMON task if needed. Point out any pumps that may have shut off*
  9) *Monitor the 1 min training on TRACK task. Correct errors if needed.*
  10) *At 3:00 mark, the RED light will come on.*
  11) *Move cursor to bottom and click on task bar to pause tasks—or click outside screen*
  12) Any questions about this task?

- **Training—System monitoring:**
  1) The system monitoring task represents an automated system such as those that operators might monitor in aircraft, rail transportation, or other environments.
  2) This task requires you to monitor gauges, which are the blue bars, and to reset the gauges if they go out of their normal ranges.
  3) You will use the F1 through F4 keys to reset the gauges. Locate the F1 through F4 keys now.
  4) The gauges are in the normal range when they fluctuate between marks in the middle of the bars. They are in the normal ranges now.
  5) Any gauge is out of the normal range when it goes to the extremes of the bar, that is when it goes to the very top or the bottom.
  6) This system is about 80% reliable in maintaining its systems in the normal ranges. Be sure to monitor the system gauges and reset them if they go out of the normal ranges.
  7) <u>Remember, this system is usually reliable. However, the main goal of this task is to monitor the gauges and to reset them when they go out of their normal ranges.</u>
  8) Let's try it. *(Restart the task by clicking inside the task area on the screen.)*
  9) *Monitor 1 min, 30 sec SYSMON task. Correct errors.*
  10) *At the 4:30 mark, the RED light will come on.*
  11) *Pause the system as before by clicking on the bottom task bar.*

- **Training—Communications:**
  1) The communications for this study occur at various times.
  2) The communications consist of a speaker giving instructions with a request for a specific response.
  3) There are two types of communications that you will hear.
     a) In one type, the speaker will say three words and you will repeat exactly what you hear.
     b) For example, the speaker may say, "Repeat, crowd, socks, puppy". You would say, "Crowd, socks, puppy."
     c) In the second type, the speaker will say one word. You will think of three words that start with the last letter of that word and say them.
     d) For example, you may hear, "Say three words starting with the last letter in Crowd". You would then think of three words that start with the letter "D" because that is the last letter in "Crowd". You might say, "Dog, dragon, day".
     e) Do you understand?
     f) Please be sure to answer each communication requests out loud.
     g) You are not being recorded. I am keeping track of communications with a checklist.
  4) Let's try this. I will start the tasks again. Continue doing all tasks as before. When a communication request occurs, just do as it says. *(Restart the task by clicking inside the task area on the screen.) If there are 2 participants, unplug the headset for only one of them.*
  5) *(Monitor the 3 min, 30 sec of the COMMS task with SYSMON events. Correct errors. The training ends at 8:00 minutes.)*

II) Completing the NASA TLX

- "At several points in the experiment, you will see this rating scale. This scale asks you about your workload in various areas. A sheet is provided next to your computer that explains each scale. For example, mental load….."
- Have the description pages available for reference here. State each scale briefly.
- **"BE SURE TO NOTE THAT THE Performance score is different from the rest. If you believe your performance was good, that rating goes to the left."**
- "Go ahead and do the workload rating scale now for practice. Wait before you save your answers.
- "Again, just remember, that Performance rating seems different from the others. If you believe your performance was good, be sure to score it to the left."
- "Go ahead and click Save All to save your ratings."
- The system will shut off after they answer the practice TLX.
  1) "Do you have any questions about any of these tasks?"
  2) "We will get started in just a few minutes. For now, please answer this brief demographics scale."
  3) Hit "NO" in last screen to close MATB-II Training file. Or hit "Yes", "File", "Exit". Data save automatically

III) Completing the Demographics form

1) Direct participant to the demographics form, or give one to them.
2) Prepare the experiment. *In "MATB: Please Select the Events File" dialogue box, select the file that corresponds to the run sheet for the study (i.e. Study1_Before) and the row (i.e. A through H). **For example**, "**MATB_Events-Study1_Before_A**".*
3) *Click "Open".*
4) *Click "File" at the top, then click "Start"*
5) *In the next window, click "Yes" or just hit "Enter" key*
6) *Move pointer to bottom of the screen to bring up the task bar*
7) *Click on the task bar to pause the MATB-II tasks*

Run the experiment
I) In the experiment you will complete several blocks where you will monitor and perform the three tasks—Resource Management, Tracking, and System Monitoring—as you did in the training.
II) Also you will hear the two types of communications. Answer each communication request as you were trained.
III) Whenever you see the Workload ratings, be sure to answer each scale and click the submit button at the bottom. The tasks will start again after the ratings.
IV) Be aware that if you do not complete the Workload ratings in 60 seconds, the system will resume. Please try to answer them promptly each time they appear.
V) If you have to stop for any reason, please let me know so that I can pause the system.
VI) The best time to pause is during the Workload rating scale.
VII) Are you ready to start?
VIII) *Click in middle of the screen to start the run.*
IX) *Hit "NO" in last screen to close MATB-II. Or hit "Yes", "File", "Exit". Data save automatically*

Finish
I) Thank-you for participating in this study.
II) *Give participant the following:*
A) *Copy of Informed Consent*
B) *Communications Educational Feedback Sheet*
C) *Departmental Participant evaluation form—Return to office on 3rd floor across from elevators*
III) *Say:* This educational feedback sheet tells you more about this study. *Explain this form and the study briefly.
IV) Here is a copy of the consent form for your records. If you have any questions or concerns, please feel free to contact one of the researchers using the information on the consent form.
V) This Evaluation form is collected by the psychology department. Please fill this out and turn in to the department office here on the 3rd floor at the end of the hallway by the elevators
VI) Feel free to leave any forms that you do not want. There is no obligation to keep any of them.
VII) *After participant leaves, update the "Dissertation Run Sheet" with any Notes or other info*
A) *Check that Sona #, date and time are in the respective columns*
B) *Enter your name in the "Notes" section*
VIII) *Check over the Participant Sheet for any errors, items left incomplete*
A) *Be sure the Sona #, Study number, row, and which participant this is in the row correspond to the run sheet.*
B) *Be sure to note which system the participant used: Dell or HP computer.*
C) *Enter any notes regarding the experiment in the notes column*
IX) *Award Credits in SONA*
X) *At the end of the day or before leaving*
A) *Close any open windows and turn off the computers and monitors*
B) *Turn off noise cancelling switch on headphones*
C) *Collect any completed forms (in case paper forms are completed if the COS system is down)*
D) *Put joystick and other materials away if needed.*
E) *Collect phone charger cord to take home*
F) *Collect all charger/power cords to take home (phone/tablet, computer)*
G) *Check to see if anything needs to be printed for the next day.*

APPENDIX D:
DEMOGRAPHICS QUESTIONNAIRE

DEMOGRAPHICS QUESTIONNAIRE

Communications in Multitask Environments

1. Age: _____________

2. Gender: Male ______ Female ______

3. Which hand do you write with? Right______ Left______

4. Is your vision at 20/20 for each eye (with or without glasses)? Yes______ No______

5. To your knowledge, are you color blind? Yes______ No______

6. Do you own or have access to a computer? Yes______ No______

7. If yes, how often do you use a computer?

   Daily_____ Several times a week ______ Occasionally_______ Never_______

8. Estimate how many hours per week you use a computer (circle one).

0-9	10-19	20-29	30-39	40+
hours	hours	hours	hours	hours

9. How do you rate your computer skills?

   Novice/Beginner______ Intermediate______ Expert_______

10. Do you use the Internet? Yes______ No______

11. Do you own or use a video game system with a joystick? Yes______ No______

12. How would you rate your video game skills?

    Novice/Beginner______ Intermediate______ Expert_______

13. What is your level of confidence with video games in general?

1	2	3	4	5
Low		Average		High

14. How many hours per week do you currently play video games?

0-9	10-19	20-29	30-39	40+
hours	hours	hours	hours	hours

15. How often do you work with aviation-related games or simulations (e.g., MS Flight Simulator, X-Plane, ProFlight Simulator, others)

    Never Rarely Monthly Weekly Daily

APPENDIX E:
COMMUNICATIONS TASK OBSERVATION SHEET

Communications Checklist

Date ______________ Participant Sona # ______________ Researcher ______________ System: ____ Dell _____HP

Circle: Study # 1 (Before) or Study #2 (After) Row Assigned _________ #Participant for row ________

(See Run Sheet, ex: 1A1 is Study 1, Row A, 1st participant for Row A)

1. Circle the Row assignment for participant on appropriate table below.
2. During the experiment, circle the words the person repeats in the "Simple" blocks
3. Write in the 3 words the participant generates for the "Complex" blocks.
4. Write in any comments in the space in the block, such as words repeated that were not the actual words or if P's says more than three words in complex box. Use the "Notes" block at the end for other notes or concerns.

Counterbalancing plans by blocks:
Study 1. "0" blocks are full pairing of communications with system monitoring malfunction events.

A	-8-S	-5-S	0-C	-2-S	-2-C	0-S	-5-C	-8-C
B	-5-S	-2-S	-8-S	0-S	0-C	-8-C	-2-C	-5-C
C	-2-S	0-S	-5-S	-8-C	-8-S	-5-C	0-C	-2-C
D	0-S	-8-C	-2-S	-5-C	-5-S	-2-C	-8-S	0-C
E	-8-C	-5-C	0-S	-2-C	-2-S	0-C	-5-S	-8-S
F	-5-C	-2-C	-8-C	0-C	0-S	-8-S	-2-S	-5-S
G	-2-C	0-C	-5-C	-8-S	-8-C	-5-S	0-S	-2-S
H	0-C	-8-S	-2-C	-5-S	-5-C	-2-S	-8-C	0-S

Block	Timing (Seconds)	Complexity
1 = -8-S	-8	Simple
2 = -5-S	-5	Simple
3 = -2-S	-2	Simple
4 = 0-S	0	Simple
5 = -8-C	-8	Complex
6 = -5-C	-5	Complex
7 = -2-C	-2	Complex
8 = 0-C	0	Complex

Study 2. "0" blocks are no pairing of communications with system monitoring malfunction events.

A	8-S	5-S	0-C	2-S	2-C	0-S	5-C	8-C
B	5-S	2-S	8-S	0-S	0-C	8-C	2-C	5-C
C	2-S	0-S	5-S	8-C	8-S	5-C	0-C	2-C
D	0-S	8-C	2-S	5-C	5-S	2-C	8-S	0-C
E	8-C	5-C	0-S	2-C	2-S	0-C	5-S	8-S
F	5-C	2-C	8-C	0-C	0-S	8-S	2-S	5-S
G	2-C	0-C	5-C	8-S	8-C	5-S	0-S	2-S
H	0-C	8-S	2-C	5-S	5-C	2-S	8-C	0-S

Block	Timing (Seconds)	Complexity
1 = 8-S	8	Simple
2 = 5-S	5	Simple
3 = 2-S	2	Simple
4 = 0-S	0-no pairing	Simple
5 = 8-C	8	Complex
6 = 5-C	5	Complex
7 = 2-C	2	Complex
8 = 0-C	0- no pairing	Complex

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	-8-S	-5-S	0-C	-2-S	-2-C	0-S	-5-C	-8-C	
	agree	banjo	various	earnest	trumps	makeup	goats	shrimp	
	burst	cancel	1	saucer	1	reply	1	1	
	manila	squeak	2	states	2	foolish	2	2	
Date:			3		3		3	3	
	vowel	skillet	earth	steady	resist	crutch	verify	blanket	
______	survey	chiefs	1	uniform	1	total	1	1	
	advisor	owner	2	except	2	curious	2	2	
Sona #			3		3		3	3	
	display	confirm	teacup	tuxedo	Cabinet	confirm	sink	impress	
______	adopt	touch	1	expert	1	poem	1	1	
	detach	serious	2	cover	2	discuss	2	2	
			3		3		3	3	
	awful	buyer		fasten	Fluid	descent	quiver	consider	
	again	thief	thesis	doing	1	plush	1	1	
Study:	employ	crisis	1	mirror	2	labor	2	2	
			2		3		3	3	
1 or 2	agony	yield	3	fortress	furious	perish	overflow	raccoon	
	legal	curfew	switch	sodium	1	knack	1	1	
	album	dismiss	1	ghetto	2	overdraw	2	2	
			2		3		3	3	
	apply	clinch	3	result	struck	punish	women	quartet	
Row A	pursue	detail	traffic	grief	1	notify	1	1	
	rapid	saucy	1	halo	2	outfit	2	2	
			2		3		3	3	
	valid	predict	3	inform	desert	prowl	establish	swift	
#for row	bonnet	defender	trunk	scissors	1	stuck	1	1	
	sleuth	venom	1	invent	2	ascend	2	2	
1 2 3			2		3		3	3	
	contact	destroy	3	janitor	adverb	spying	crept	wisdom	
4 5 6	moist	arrow	concern	brown	1	public	1	1	
	verbal	syntax	1	limit	2	emotion	2	2	
			2		3		3	3	
System:	poultry	packet	3	defeat	supply	admit	solo	echo	
	crystal	assign	vary	algebra	1	referee	1	1	
Dell	cement	duet	1	knob	2	mislead	2	2	
			2		3		3	3	
HP	basin	refer	3	lungs	zero	relief	detain	instead	
	copy	deposit	taunt	import	1	common	1	1	
	soprano	giant	1	salmon	2	rumor	2	2	
			2		3		3	3	
			3						

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	-5-S	-2-S	-8-S	0-S	0-C	-8-C	-2-C	-5-C	
Date: ________	banjo cancel squeak	earnest saucer states	agree burst manila	makeup reply foolish	various 1 2 3	shrimp 1 2 3	trumps 1 2 3	goats 1 2 3	
Sona #	skillet chiefs owner	steady uniform except	vowel survey advisor	crutch total curious	earth 1 2 3	blanket 1 2 3	resist 1 2 3	verify 1 2 3	
________	confirm touch serious	tuxedo expert cover	display adopt detach	confirm poem discuss	teacup 1 2 3	impress 1 2 3	Cabinet 1 2 3	sink 1 2 3	
Study:	buyer thief crisis	fasten doing mirror	awful again employ	descent plush labor	thesis 1 2 3	consider 1 2 3	Fluid 1 2 3	quiver 1 2 3	
1 or 2	yield curfew dismiss	fortress sodium ghetto	agony legal album	perish knack overdraw	switch 1 2 3	raccoon 1 2 3	furious 1 2 3	overflow 1 2 3	
Row B	clinch detail saucy	result grief halo	apply pursue rapid	punish notify outfit	traffic 1 2 3	quartet 1 2 3	struck 1 2 3	women 1 2 3	
#for row 1 2 3	predict defender venom	inform scissors invent	valid bonnet sleuth	prowl stuck ascend	trunk 1 2 3	swift 1 2 3	desert 1 2 3	establish 1 2 3	
4 5 6	destroy arrow syntax	janitor brown limit	contact moist verbal	spying public emotion	concern 1 2 3	wisdom 1 2 3	adverb 1 2 3	crept 1 2 3	
System: Dell	packet assign duet	defeat algebra knob	poultry crystal cement	admit referee mislead	vary 1 2 3	echo 1 2 3	supply 1 2 3	solo 1 2 3	
HP	refer deposit giant	lungs import salmon	basin copy soprano	relief common rumor	taunt 1 2 3	instead 1 2 3	zero 1 2 3	detain 1 2 3	

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	-2-S	0-S	-5-S	--8-C	-8-S	--5-C	0-C	-2-C	
	earnest	makeup	banjo	shrimp	agree	goats	various	trumps	
Date:	saucer	reply	cancel	1	burst	1	1	1	
	states	foolish	squeak	2	manila	2	2	2	
				3		3	3	3	
________	steady	crutch	skillet	blanket	vowel	verify	earth	resist	
	uniform	total	chiefs	1	survey	1	1	1	
Sona #	except	curious	owner	2	advisor	2	2	2	
				3		3	3	3	
	tuxedo	confirm	confirm	impress	display	sink	teacup	Cabinet	
________	expert	poem	touch	1	adopt	1	1	1	
	cover	discuss	serious	2	detach	2	2	2	
				3		3	3	3	
	fasten	descent	buyer	consider	awful	quiver	thesis	Fluid	
	doing	plush	thief	1	again	1	1	1	
Study:	mirror	labor	crisis	2	employ	2	2	2	
				3		3	3	3	
	fortress	perish	yield	raccoon	agony	overflow	switch	furious	
1 or 2	sodium	knack	curfew	1	legal	1	1	1	
	ghetto	overdraw	dismiss	2	album	2	2	2	
				3		3	3	3	
	result	punish	clinch	quartet	apply	women	traffic	struck	
Row C	grief	notify	detail	1	pursue	1	1	1	
	halo	outfit	saucy	2	rapid	2	2	2	
				3		3	3	3	
	inform	prowl	predict	swift	valid	establish	trunk	desert	
#for row	scissors	stuck	defender	1	bonnet	1	1	1	
	invent	ascend	venom	2	sleuth	2	2	2	
1 2 3				3		3	3	3	
	janitor	spying	destroy	wisdom	contact	crept	concern	adverb	
4 5 6	brown	public	arrow	1	moist	1	1	1	
	limit	emotion	syntax	2	verbal	2	2	2	
				3		3	3	3	
System:	defeat	admit	packet	echo	poultry	solo	vary	supply	
	algebra	referee	assign	1	crystal	1	1	1	
Dell	knob	mislead	duet	2	cement	2	2	2	
				3		3	3	3	
HP	lungs	relief	refer	instead	basin	detain	taunt	zero	
	import	common	deposit	1	copy	1	1	1	
	salmon	rumor	giant	2	soprano	2	2	2	
				3		3	3	3	

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	0-S	-8-C	-2-S	-5-C	-5-S	-2-C	-8-S	0-C	
Date:	makeup reply foolish	shrimp 1 2 3	earnest saucer states	goats 1 2 3	banjo cancel squeak	trumps 1 2 3	agree burst manila	various 1 2 3	
______ Sona #	crutch total curious	blanket 1 2 3	steady uniform except	verify 1 2 3	skillet chiefs owner	resist 1 2 3	vowel survey advisor	earth 1 2 3	
______	confirm poem discuss	impress 1 2 3	tuxedo expert cover	sink 1 2 3	confirm touch serious	Cabinet 1 2 3	display adopt detach	teacup 1 2 3	
Study:	descent plush labor	consider 1 2 3	fasten doing mirror	quiver 1 2 3	buyer thief crisis	Fluid 1 2 3	awful again employ	thesis 1 2 3	
1 or 2	perish knack overdraw	raccoon 1 2 3	fortress sodium ghetto	overflow 1 2 3	yield curfew dismiss	furious 1 2 3	agony legal album	switch 1 2 3	
Row D	punish notify outfit	quartet 1 2 3	result grief halo	women 1 2 3	clinch detail saucy	struck 1 2 3	apply pursue rapid	traffic 1 2 3	
#for row 1 2 3	prowl stuck ascend	swift 1 2 3	inform scissors invent	establish 1 2 3	predict defender venom	desert 1 2 3	valid bonnet sleuth	trunk 1 2 3	
4 5 6	spying public emotion	wisdom 1 2 3	janitor brown limit	crept 1 2 3	destroy arrow syntax	adverb 1 2 3	contact moist verbal	concern 1 2 3	
System: Dell	admit referee mislead	echo 1 2 3	defeat algebra knob	solo 1 2 3	packet assign duet	supply 1 2 3	poultry crystal cement	vary 1 2 3	
HP	relief common rumor	instead 1 2 3	lungs import salmon	detain 1 2 3	refer deposit giant	zero 1 2 3	basin copy soprano	taunt 1 2 3	

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	-8-C	-5-C	0-S	-2-C	-2-S	0-C	-5-S	-8-S	
Date:	shrimp 1 2 3	goats 1 2 3	makeup reply foolish	trumps 1 2 3	earnest saucer states	various 1 2 3	banjo cancel squeak	agree burst manila	
________ Sona #	blanket 1 2 3	verify 1 2 3	crutch total curious	resist 1 2 3	steady uniform except	earth 1 2 3	skillet chiefs owner	vowel survey advisor	
________	impress 1 2 3	sink 1 2 3	confirm poem discuss	Cabinet 1 2 3	tuxedo expert cover	teacup 1 2 3	confirm touch serious	display adopt detach	
Study:	consider 1 2 3	quiver 1 2 3	descent plush labor	Fluid 1 2 3	fasten doing mirror	thesis 1 2 3	buyer thief crisis	awful again employ	
1 or 2	raccoon 1 2 3	overflow 1 2 3	perish knack overdraw	furious 1 2 3	fortress sodium ghetto	switch 1 2 3	yield curfew dismiss	agony legal album	
Row E	quartet 1 2 3	women 1 2 3	punish notify outfit	struck 1 2 3	result grief halo	traffic 1 2 3	clinch detail saucy	apply pursue rapid	
#for row 1 2 3	swift 1 2 3	establish 1 2 3	prowl stuck ascend	desert 1 2 3	inform scissors invent	trunk 1 2 3	predict defender venom	valid bonnet sleuth	
4 5 6	wisdom 1 2 3	crept 1 2 3	spying public emotion	adverb 1 2 3	janitor brown limit	concern 1 2 3	destroy arrow syntax	contact moist verbal	
System: Dell	echo 1 2 3	solo 1 2 3	admit referee mislead	supply 1 2 3	defeat algebra knob	vary 1 2 3	packet assign duet	poultry crystal cement	
HP	instead 1 2 3	detain 1 2 3	relief common rumor	zero 1 2 3	lungs import salmon	taunt 1 2 3	refer deposit giant	basin copy soprano	

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	-5-C	-2-C	-8-C	0-C	0-S	-8-S	-2-S	-5-S	
Date: ________	goats 1 2 3	trumps 1 2 3	shrimp 1 2 3	various 1 2 3	makeup reply foolish	agree burst manila	earnest saucer states	banjo cancel squeak	
Sona # ________	verify 1 2 3	resist 1 2 3	blanket 1 2 3	earth 1 2 3	crutch total curious	vowel survey advisor	steady uniform except	skillet chiefs owner	
	sink 1 2 3	Cabinet 1 2 3	impress 1 2 3	teacup 1 2 3	confirm poem discuss	display adopt detach	tuxedo expert cover	confirm touch serious	
Study:	quiver 1 2 3	Fluid 1 2 3	consider 1 2 3	thesis 1 2 3	descent plush labor	awful again employ	fasten doing mirror	buyer thief crisis	
1 or 2	overflow 1 2 3	furious 1 2 3	raccoon 1 2 3	switch 1 2 3	perish knack overdraw	agony legal album	fortress sodium ghetto	yield curfew dismiss	
Row F	women 1 2 3	struck 1 2 3	quartet 1 2 3	traffic 1 2 3	punish notify outfit	apply pursue rapid	result grief halo	clinch detail saucy	
#for row 1 2 3	establish 1 2 3	desert 1 2 3	swift 1 2 3	trunk 1 2 3	prowl stuck ascend	valid bonnet sleuth	inform scissors invent	predict defender venom	
4 5 6	crept 1 2 3	adverb 1 2 3	wisdom 1 2 3	concern 1 2 3	spying public emotion	contact moist verbal	janitor brown limit	destroy arrow syntax	
System: Dell	solo 1 2 3	supply 1 2 3	echo 1 2 3	vary 1 2 3	admit referee mislead	poultry crystal cement	defeat algebra knob	packet assign duet	
HP	detain 1 2 3	zero 1 2 3	instead 1 2 3	taunt 1 2 3	relief common rumor	basin copy soprano	lungs import salmon	refer deposit giant	

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	--2-C	0-C	-5-C	-8-S	-8-C	-5-S	0-S	-2-S	
Date:	trumps 1 2 3	various 1 2 3	goats 1 2 3	agree burst manila	shrimp 1 2 3	banjo cancel squeak	makeup reply foolish	earnest saucer states	
________ Sona #	resist 1 2 3	earth 1 2 3	verify 1 2 3	vowel survey advisor	blanket 1 2 3	skillet chiefs owner	crutch total curious	steady uniform except	
________	Cabinet 1 2 3	teacup 1 2 3	sink 1 2 3	display adopt detach	impress 1 2 3	confirm touch serious	confirm poem discuss	tuxedo expert cover	
Study:	Fluid 1 2 3	thesis 1 2 3	quiver 1 2 3	awful again employ	consider 1 2 3	buyer thief crisis	descent plush labor	fasten doing mirror	
1 or 2	furious 1 2 3	switch 1 2 3	overflow 1 2 3	agony legal album	raccoon 1 2 3	yield curfew dismiss	perish knack overdraw	fortress sodium ghetto	
Row G	struck 1 2 3	traffic 1 2 3	women 1 2 3	apply pursue rapid	quartet 1 2 3	clinch detail saucy	punish notify outfit	result grief halo	
#for row 1 2 3	desert 1 2 3	trunk 1 2 3	establish 1 2 3	valid bonnet sleuth	swift 1 2 3	predict defender venom	prowl stuck ascend	inform scissors invent	
4 5 6	adverb 1 2 3	concern 1 2 3	crept 1 2 3	contact moist verbal	wisdom 1 2 3	destroy arrow syntax	spying public emotion	janitor brown limit	
System: Dell	supply 1 2 3	vary 1 2 3	solo 1 2 3	poultry crystal cement	echo 1 2 3	packet assign duet	admit referee mislead	defeat algebra knob	
HP	zero 1 2 3	taunt 1 2 3	detain 1 2 3	basin copy soprano	instead 1 2 3	refer deposit giant	relief common rumor	lungs import salmon	

Row	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8	Notes
	0-C	-8-S	-2-C	-5-S	-5-C	-2-S	-8-C	0-S	
	various	agree	trumps	banjo	goats	earnest	shrimp	makeup	
	1	burst	1	cancel	1	saucer	1	reply	
Date:	2	manila	2	squeak	2	states	2	foolish	
	3		3		3		3		
________	earth	vowel	resist	skillet	verify	steady	blanket	crutch	
	1	survey	1	chiefs	1	uniform	1	total	
	2	advisor	2	owner	2	except	2	curious	
Sona #	3		3		3		3		
	teacup	display	Cabinet	confirm	sink	tuxedo	impress	confirm	
	1	adopt	1	touch	1	expert	1	poem	
________	2	detach	2	serious	2	cover	2	discuss	
	3		3		3		3		
	thesis	awful	Fluid	buyer	quiver	fasten	consider	descent	
	1	again	1	thief	1	doing	1	plush	
Study:	2	employ	2	crisis	2	mirror	2	labor	
	3		3		3		3		
	switch	agony	furious	yield	overflow	fortress	raccoon	perish	
1 or 2	1	legal	1	curfew	1	sodium	1	knack	
	2	album	2	dismiss	2	ghetto	2	overdraw	
	3		3		3		3		
	traffic	apply	struck	clinch	women	result	quartet	punish	
Row H	1	pursue	1	detail	1	grief	1	notify	
	2	rapid	2	saucy	2	halo	2	outfit	
	3		3		3		3		
	trunk	valid	desert	predict	establish	inform	swift	prowl	
#for row	1	bonnet	1	defender	1	scissors	1	stuck	
	2	sleuth	2	venom	2	invent	2	ascend	
1 2 3	3		3		3		3		
	concern	contact	adverb	destroy	crept	janitor	wisdom	spying	
4 5 6	1	moist	1	arrow	1	brown	1	public	
	2	verbal	2	syntax	2	limit	2	emotion	
	3		3		3		3		
System:	vary	poultry	supply	packet	solo	defeat	echo	admit	
	1	crystal	1	assign	1	algebra	1	referee	
Dell	2	cement	2	duet	2	knob	2	mislead	
	3		3		3		3		
HP	taunt	basin	zero	refer	detain	lungs	instead	relief	
	1	copy	1	deposit	1	import	1	common	
	2	soprano	2	giant	2	salmon	2	rumor	
	3		3		3		3		

APPENDIX F:
SYSTEM MONITORING DATA FILE EXAMPLE

```
# 04-11-2014     09:30:31     SYSM_2014_04110930.txt
#
# Events Filename: MATB_EVENTS-Study1_Before_G.xml
#
# Timeout (in seconds):  Lights = 30   Scales = 30
#
# RT = Response Time (in seconds)
# SYS_OK = An event for the system selected is active
#
#-TIME-      -RT-    -SYSTEM- -LIGHT/SCALE-  -SYS_OK-  -REMARKS-
#--------------------------------------------------------------------
 00:00:44.0   04.0    Scale    ONE            TRUE
 00:01:19.0   05.0    Scale    TWO            TRUE
 00:01:51.2   04.2    Scale    FOUR           TRUE
 00:02:49.0   04.0    Scale    TWO            TRUE
 00:03:15.3   02.3    Scale    THREE          TRUE
 00:03:50.8   03.8    Scale    THREE          TRUE
 00:04:28.9   08.9    Scale    FOUR           TRUE
 00:04:48.5   03.5    Scale    THREE          TRUE
 00:05:43.4   05.4    Scale    ONE            TRUE
 00:06:41.9  -10      Scale    FOUR                  - Event Timedout
 00:07:45.4  -10      Scale    THREE                 - Event Timedout
 00:07:46.8           Scale    THREE          FALSE
 00:08:31.5   04.5    Scale    ONE            TRUE
 00:10:02.5  -10      Scale    TWO                   - Event Timedout
 00:10:04.2           Scale    TWO            FALSE
 00:10:19.6   02.6    Scale    ONE            TRUE
 00:10:46.4   04.4    Scale    FOUR           TRUE
 00:11:22.4  -10      Scale    TWO                   - Event Timedout
 00:11:23.1           Scale    TWO            FALSE
 00:11:39.2   02.2    Scale    FOUR           TRUE
 00:12:25.3   03.3    Scale    TWO            TRUE
 00:12:54.8   07.8    Scale    THREE          TRUE
 00:13:15.0   03.0    Scale    ONE            TRUE
 00:14:37.0   02.0    Scale    TWO            TRUE
 00:15:20.9  -10      Scale    ONE                   - Event Timedout
 00:16:16.6   04.6    Scale    THREE          TRUE
 00:17:08.0   06.0    Scale    THREE          TRUE
 00:17:44.5   05.5    Scale    FOUR           TRUE
 00:18:15.9   02.9    Scale    TWO            TRUE
 00:18:51.0   03.0    Scale    ONE            TRUE
 00:19:25.4   07.4    Scale    FOUR           TRUE
 00:20:05.0  -10      Scale    THREE                 - Event Timedout
 00:20:05.3           Scale    THREE          FALSE
 00:20:49.9  -10      Scale    ONE                   - Event Timedout
 00:20:50.2           Scale    ONE            FALSE
 00:21:29.0   04.0    Scale    THREE          TRUE
 00:22:05.1   02.1    Scale    ONE            TRUE
 00:22:28.6   02.6    Scale    FOUR           TRUE
 00:23:28.9   03.9    Scale    TWO            TRUE
 00:24:08.7   03.7    Scale    FOUR           TRUE
 00:25:17.5   04.5    Scale    ONE            TRUE
```

00:26:13.3	09.3	Scale	TWO	TRUE	
00:26:47.1	02.1	Scale	THREE	TRUE	
00:27:27.8	02.8	Scale	TWO	TRUE	
00:27:43.2	01.2	Scale	FOUR	TRUE	
00:28:30.6		Scale	TWO	FALSE	
00:28:34.9	-10	Scale	THREE		- Event Timedout
00:29:05.8	02.8	Scale	ONE	TRUE	
00:29:36.0	-10	Scale	FOUR		- Event Timedout
00:30:34.9	-10	Scale	TWO		- Event Timedout
00:31:09.3	04.3	Scale	FOUR	TRUE	
00:32:17.0	04.0	Scale	ONE	TRUE	
00:33:06.3	02.3	Scale	TWO	TRUE	
00:33:52.6	07.6	Scale	THREE	TRUE	
00:34:35.0	-10	Scale	TWO		- Event Timedout
00:34:46.2	04.2	Scale	FOUR	TRUE	
00:35:37.4	02.4	Scale	TWO	TRUE	
00:36:13.5	02.5	Scale	ONE	TRUE	
00:37:15.4	03.4	Scale	THREE	TRUE	
00:38:06.8	04.8	Scale	THREE	TRUE	
00:38:41.8	02.8	Scale	FOUR	TRUE	
00:39:15.8	02.8	Scale	TWO	TRUE	
00:39:51.2	03.2	Scale	ONE	TRUE	
00:40:19.2	01.2	Scale	FOUR	TRUE	
00:40:58.3	03.3	Scale	THREE	TRUE	
00:41:49.9	09.9	Scale	ONE	TRUE	
00:42:37.9	02.9	Scale	THREE	TRUE	
00:43:34.0	07.0	Scale	ONE	TRUE	
00:44:59.9	07.9	Scale	TWO	TRUE	
00:45:20.8	03.8	Scale	ONE	TRUE	
00:45:44.3	02.3	Scale	FOUR	TRUE	
00:46:19.5	07.5	Scale	TWO	TRUE	
00:46:43.0	06.0	Scale	FOUR	TRUE	
00:47:24.5	02.5	Scale	TWO	TRUE	
00:47:54.4	07.4	Scale	THREE	TRUE	
00:48:14.7	02.7	Scale	ONE	TRUE	
00:49:50.0	-10	Scale	ONE		- Event Timedout
00:50:20.4	06.4	Scale	TWO	TRUE	
00:50:52.8	05.8	Scale	FOUR	TRUE	
00:51:48.8	03.8	Scale	TWO	TRUE	
00:52:18.4	05.4	Scale	THREE	TRUE	
00:52:49.3	02.3	Scale	THREE	TRUE	
00:53:22.0	02.0	Scale	FOUR	TRUE	
00:53:48.1	03.1	Scale	THREE	TRUE	
00:54:42.3	04.3	Scale	ONE	TRUE	
00:55:35.8	03.8	Scale	FOUR	TRUE	

APPENDIX G:
Z-SCORE MEANS AND VARIABLITY ESTIMATES
FOR PERFORMANCE MEASURES IN FIGURES OF MERIT

Performance Measures	Communication Complexity		Communication Timing Intervals Relative to Visual Detection Events -8	-5	-2	0	2	5	8	15+
Accuracy	Simple	Mean z-score	0.00	0.06	-0.04	0.03	0.26	0.30	0.16	0.26
		SE	0.14	0.19	0.22	0.19	0.16	0.14	0.23	0.11
		SD	0.69	0.92	1.06	0.94	0.79	0.67	1.13	0.53
	Complex	Mean z-score	-0.14	-0.31	-0.17	-1.11	0.23	0.20	0.06	0.20
		SE	0.18	0.24	0.29	0.30	0.13	0.13	0.21	0.17
		SD	0.91	1.17	1.41	1.45	0.63	0.62	1.04	0.82
Response time	Simple	Mean z-score	0.15	-0.06	-0.12	0.25	-0.30	-0.39	-0.33	-0.21
		SE	0.15	0.21	0.20	0.18	0.18	0.19	0.26	0.15
		SD	0.72	1.03	0.98	0.90	0.88	0.94	1.26	0.75
	Complex	Mean z-score	0.02	0.28	0.54	0.86	-0.10	-0.14	-0.24	-0.22
		SE	0.19	0.23	0.22	0.24	0.15	0.16	0.23	0.16
		SD	0.93	1.12	1.08	1.18	0.75	0.79	1.12	0.78
Tracking	Simple	Mean z-score	-0.08	0.02	0.13	-0.02	-0.07	-0.05	-0.01	0.03
		SE	0.22	0.22	0.17	0.17	0.19	0.22	0.20	0.20
		SD	1.06	1.07	0.86	0.84	0.94	1.06	1.00	0.96
	Complex	Mean z-score	0.26	0.09	-0.03	-0.08	-0.04	-0.07	-0.06	-0.03
		SE	0.24	0.23	0.22	0.19	0.20	0.19	0.21	0.23
		SD	1.20	1.12	1.07	0.95	0.96	0.94	1.03	1.15
Workload	Simple	Mean z-score	0.05	0.06	-0.30	-0.43	0.05	0.06	0.07	0.06
		SE	0.19	0.20	0.24	0.22	0.19	0.18	0.18	0.20
		SD	0.92	0.99	1.17	1.08	0.92	0.87	0.91	0.99
	Complex	Mean z-score	0.48	0.39	-0.05	0.01	0.48	0.36	0.28	0.39
		SE	0.16	0.16	0.21	0.23	0.16	0.16	0.17	0.16
		SD	0.79	0.78	1.03	1.14	0.79	0.79	0.83	0.78

Notes: 1) Values of -8 to 0 were obtained from study 1. Values from 2 through 15+ were obtained from study 2.
2) The "0" conditions were communications presented simultaneously with visual detection events.

REFERENCES

Airbus and Flight Safety Foundation. (2004a). Flight operations briefing notes: Effective pilot/controller conversation. [Online industry document]. Retrieved from http://www.airbus.com/fileadmin/media_gallery/files/safety_library_items/AirbusSafetyLib_-FLT_OPS-HUM_PER-SEQ04.pdf

Airbus and Flight Safety Foundation. (2004b). Flight operations briefing notes: Managing interruptions and distractions. [Online industry document]. Retrieved from http://www.airbus.com/fileadmin/media_gallery/files/safety_library_items/AirbusSafetyLib_-FLT_OPS-HUM_PER-SEQ03.pdf

Angell, L.S., Auflick, J., Austria, P. A., Kochhar, D.,Tijerina, L. Biever, W., Diptiman, T., Hogsett, J., & Kiger, S. (2006). *Driver workload metrics Task 2 final report and appendices* (Report No. DOT HS 810 635). Washington, DC: National Highway Traffic Safety Administration. Retrieved from http://trid.trb.org/view.aspx?id=855927

Baddeley, A.D. (1986). *Working Memory.* Oxford: Oxford University Press.

Baddeley, A.D. (2003). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience, 4,* 829-839.

Bahrick, H. P., Noble, M., & Fitts, P. M. (1954). Extra task performance as a measure of learning a primary task. *Journal of Experimental Psychology, 48,* 298–302.

Bahrick, H. P., & Shelly, C. (1958). Time sharing as an index of automization. *Journal of Experimental Psychology, 56,* 288–293.

Briggs, G., Peters, G., & Fisher, R. (1972). On the locus of the divided attention effects. *Perception & Psychophysics, 11,* 315–320.

Bürki-Cohen , J. (1995). Say again? How complexity and format of air traffic control instructions affect pilot recall. In B. J. Kanki & O. V. Prinzo (Eds.), *Methods and metrics of voice conversation* (Office of Aviation Medicine Tech. Rep. No. DOT/FAA/AM-96/10). Retrieved from https://www.hf.faa.gov/docs/508/docs/volpeBurki1995.pdf

Cain, B. (2004). *A review of the mental workload literature* (Defence Research and Development Canada, Report No. RTO-TR-HFM-121-Part-II). Retrieved from http://ftp.rta.nato.int/public/pubfulltext/rto/tr/rto-tr-hfm-121-part-ii/tr-hfm-121-part-ii-04.pdf

Cardosi, K. M. (1993). *An analysis of en route controller-pilot voice conversation* (Office of Research and Development Rep. No. DOT/FAA/RD-93/11). Washington, DC: U.S. Department of Transportation.

Cardosi, K.M., Brett, B., & Han, S. (1996). *An analysis of TRACON (Terminal Radar Approach Control) controller-pilot voice conversation.* (Office of Research and Development Rep. No. DOT/FAA/AR-96/66). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a313942.pdf

Cardosi, K.M., Falzarano, B.B., & Han, S. (1998). *Pilot-controller conversation errors: An analysis of Aviation Safety Reporting System (ASRS) reports.* (DOT/FAA/AR-98/17). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a356655.pdf

Chattratichart, J. & Lindgaard, G. (2009). Is the 'Figure of Merit' Really That Meritorious? *Human-Computer Interaction–INTERACT 2009*, (pp. 235-238). Berlin: Springer.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Connell, L. (1995). Pilot and controller conversation issues. In B.G. Kanki & C.W. Prinzo (Eds.), *Proceedings of the Methods and Metrics of Voice Conversation Workshop,* (DOT/FAA/AM-96/10). Washington, D.C.: Office of Aviation Medicine.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87-185.

Cushing, S. (1994). Fatal words: conversation clashes and aircraft crashes. Chicago, IL: *The University of Chicago Press.*

Damos, D. (1997). Using interruptions to identify task prioritization in Part 121 air carrier operations. In R. S. Jensen (Ed.), Proceedings of the 9th International Symposium on Aviation Psychology (pp. 871–876). Columbus, OH: Department of Aerospace Engineering, Ohio State University.

Dismukes, K. (2001). The challenge of managing interruptions, distractions, and deferred tasks. In *Proceedings of the 11th International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University.

Damos, D.L. & Tabachnick, B.G. (2001). The effect of interruptions on flight crew performance: ASRS reports. Los Angeles, CA: Damos Research Associates.

Dismukes, R. K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airlines accidents*. Burlington, VT: Ashgate.

Dismukes, R. K., Loukopoulos, L.D., & Jobe, K. (2001). The challenge of managing interruptions, distractions, and deferred tasks. In R. Jensen (ed.), *Proceedings of the 11th International symposium on aviation psychology*. Columbus: Ohio State University.

Einstein, G.O., & McDaniel, M.A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 717–726.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191

Federal Aviation Administration. (2013a).Data Conversation (Data Comm). Retrieved from http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/atc_comms_services/datacomm

Federal Aviation Administration. (2013b). System-Wide Information Management (SWIM). Retrieved from http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/atc_comms_services/swim

Flach, J.M., & Kuperman, G. (2001). The human capacity for work: A (biased) historical perspective. In P.A. Hancock & P.A. Desmond (2001). *Stress, workload, and fatigue,* (pp. 429-442). Mahwah, NJ: Lawrence Erlbaum Associates.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 501–518.

Graf, P., & Uttl, B. (2001). Prospective memory: A new focus for research. *Consciousness and Cognition*, *10*(4), 437-450.

Hancock, P.A. (2011). *Cognitive differences in the ways men and women perceive the dimension and duration of time.* Lewiston, New York: The Edwin Mellen Press.

Hart, S.G. & Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*, (pp. 239-250). Amsterdam: North-Holland Press.

Hartson, H.R., Andre, T.S., & Williges, R.C. (2003). Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction, 15*(1), 145-181.

Harvey, C.M., Reynolds, M., Pacley, A.L., Koubek, R.J., & Rehmann, A.J. (1992). Effects of the controller-to-pilot data link (DataLink) on crew conversation. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, 61-65.

Helmreich, R.L. (1997) Managing human error in aviation. *Scientific American. 276*, (5), 62-67.

Helleberg, J.R., & Wickens, C.D. (2003). Effects of data-link modality and display redundancy on pilot performance: An attentional perspective. *The International Journal of Aviation Psychology, 13,* 189-210.

Horrey, W.J. (2011). Assessing the effects of in-vehicle tasks on driving performance. *Ergonomics in Design: The Quarterly of Human Factors Applications, 11,* 4-7.

Horrey, W., & Wickens, C.D. (2004). Driving and side task performance. *Human Factors, 46,* 611-624.

Horrey, W.J., & Wickens, C.D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48,* 196-205.

Iani, C. & Wickens, C.D. (2007). Factors affecting task management in aviation. *Human Factors, 49,* 16-24.

Jex, H.R. (1988). Measuring mental workload: Problems, progress and promises. In P.A. Hancock & N. Meshkati (Eds), *Human Mental Workload,* (pp. 5-36). Amsterdam: North-Holland.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Kalsbeek, J.W.H. & Sykes, R.N. (1967). Objective measurement of mental workload. *Acta Psychologica, 27,* 253-261.

Kanki, B. G., & Palmer, M. T. (1993). Conversation and crew, resource management. In E. Wiener, B. Kanki, & R. Helmreich (Eds.), *Cockpit resource management* (pp. 99-136). San Diego, CA: Academic Press.

Krivonos, P.D. (2007, June). *Conversation in aviation safety: Lessons learned and lessons required.* Paper presented at the 2007 Regional Seminar of the Australia and New Zealand Societies of Air Safety Investigators. Retrieved from http://www.asasi.org/papers/2007/Conversation_in_Aviation_Safety_Paul_Krivonos.pdf

Latorella, K. A. (1996). Investigating interruptions: An example from the flightdeck. In Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society (pp. 249–253). Santa Monica, CA: Human Factors and Ergonomics Society.

Latorella, K. A. (1998). Effects of modality on interrupted flight deck performance: Implications for data link. In Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (pp. 87–91). Santa Monica, CA: Human Factors and Ergonomics Society.

Latorella, K. A. (1999). Investigating interruptions: Implications for flightdeck performance (Tech. Rep. NASA/TM-1999-209707). Hampton, VA: NASA Langley Research Center.

Lee, S. J. (2011). Figures of merit in engineering design [Website post]. Retrieved from http://www.engr.sjsu.edu/sjlee/figure-of-merit.htm

Lee, J.D., & See, K.A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors, 46*, 50–80.

LeMay, M. & Comstock, R. (1990). An initial test of a normative figure of merit for the quality of overall task performance. *Proceedings of the Human Factors and Ergonomics Society 34th Annual Meeting,* 81-85.

Lesch, M.F., & Hancock, P.A. (2004). Driving performance during concurrent cellphone use: Are drivers aware of their performance decrements? *Accident Analysis and Prevention, 36,* 471-480.

Levy, J., & Pashler, H. (2008). Task prioritization in multitasking during driving: Opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology, 22,* 5-7-525.

Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 49*(3), 376-399.

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2001). Cockpit interruptions and distractions: A line observation study. *Proceedings of the 11th International Symposium on Aviation Psychology.* Dayton, OH.

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2003). Concurrent task demands in the cockpit: Challenges and vulnerabilities in routine flight operations. *Proceedings of the 12th International Symposium on Aviation Psychology.* Dayton, OH. Retrieved from http://humansystems.arc.nasa.gov/flightcognition/Publications/LL_KD_IB_ISAP_03.pdf

Loukopoulos, L.D., Dismukes, R.K., & Barshi, I. (2009). *The Multitasking Myth: Handling Complexity in Real-World Operations.* United Kingdom: Ashgate.

Lu, S.A., Wickens, C.D., Prinet, J.C., Hutchins, S.D., Sarter, N., & Sebok, A. (2013). Supporting interruption management and multimodal interface design: Three meta-analyses of task performance as a function of interrupting task modality. *Human Factors, 55*, 697-724.

Lu, Z.-L., Williamson, S.J. & Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science, 258*(5088), 1668-1670.

Maccoby, E.E. & Jacklin, C.N. (1974). *The psychology of sex differences.* Stanford, California: Stanford University Press.

McFarlane, D.C., & Latorella, K.A. (2002). The score and importance of human interruptions in human-computer interaction design. *Human-Computer Interaction, 17,* 1-61.

Meshkati, N., Hancock, P.A., Rahimi, M. & Dawes, S.M. (1995). Techniques in mental workload assessment. In: J. Wilson and E.N. Corlett, (Eds.), *Evaluation of human work: A practical ergonomics methodology, 2nd Edition,* (pp. 749-782). London: Taylor and Francis.

McGann, A., Morrow, D., Rodvold, M., & Mackintosh, M-A. (2009). Mixed-media conversation on the flight deck: A comparison of voice, data link, and mixed ATC environments. *The International Journal of Aviation Psychology,83,* 137-156.

Miyake, S., Loslever, P., & Hancock, P.A. (2001). Individual differences in tracking. *Ergonomics, 44,* 1056-1068.

Monan, W.P. (1983). *Addressee errors in ATC conversation: The call sign problem.* (NASA Contractor Report 166462). Moffett Field, CA: National Aeronautics and Space Administration.

Monan, W.P. (1988). *Human factors in air-carrier operations: The hearback problem*. (NASA Report CR 177398). Moffett Field, CA: National Aeronautics and Space Administration.

Monk, C.A., Trafton, J.G. & Boehm-Davis, D.A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied, 14,* 299-313.

Moray, N. (1967). Where is attention limited? A survey and a model. *Acta Psychologica, 27,* 84-92.

Moray, N. (1979). *Mental workload: Its theory and measurement.* New York: Plenum.

Morrow, D., & Rodvold, M. (1993). The influence of ATC message length and timing on pilot conversation (NASA Contractor Rep. No. 177621). Moffett Field, CA: NASA-Ames Research Center.

Morrow, D., Rodvold, M., & Lee, A. (1994). Nonroutine transactions in controller-pilot conversation. *Discourse Processes, 17*, 235–258.

Mosier, K.L., Rettenmaier, P., McDearmid, M., Wilson, J., Mak, S., Raj, L., & Orasanu, J. (2010). Pilot/ATC conversation conflicts: Implications for NextGen. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54,* 1521-1525.

Mouloua, M., Hancock, P. A., Jones, L. M., & Vincenzi, D. (2010). Automation in aviation systems: Issues and considerations. In J. Wise, V. D. Hopkin and D. J. Garland (Ed.), *Handbook of Aviation Human Factors, 2nd Ed,* (pp. 8-1 – 8-11). CRC Press.

Mouloua, M., Parasuraman, R., & Molloy, R. (1993). Monitoring automation failures: Effects of single and multi-adaptive function allocation. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting,* 1-5.

National Aeronautics and Space Administration, Langley Research Center (NASA-LARC) (2014). MATB-II Official Website. Retrieved from http://matb.larc.nasa.gov/

National Transportation Safety Board (NTSB) (2004). Transportation Safety Most Wanted List. Retrieved from https://www.ntsb.gov/safety/mwl.html

Neely, A. (2007). *Business performance measurement: Unifying theories and integrating practice*, (2nd ed.). Cambridge, UK: Cambridge University Press. Retrieved from http://www.alnap.org/resource/10432

Neider, M. B., McCarley, J. S., Crowell, J. A., Kaczmarski, H., & Kramer, A. F. (2010). Pedestrians, vehicles, and cell phones. *Accident Analysis & Prevention*, *42*(2), 589-594.

Neider, M. B., Gaspar, J. G., McCarley, J. S., Crowell, J. A., Kaczmarski, H., & Kramer, A. F. (2011). Walking and talking: dual-task effects on street crossing behavior in older adults. *Psychology and aging*, *26*(2), 260.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, *7(4)*, 480-494.

Orasanu, J., Martin, L., & Davison, J. (2002). Cognitive and contextual factors in aviation accidents: Decision errors. In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 209–225). Mahwah, NJ: Lawrence Erlbaum.

Pallant, J. (2007). *SPSS Survival Manual* (3rd Ed). New York, NY: Open University Press, McGraw Hill Education.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*(4), 665-679.

Parasuraman, R. & Manzey, D.H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52,* 381-410.

Prinzo, O. V. (1996). *An analysis of approach control/pilot voice conversation.* (Office of Aviation Medicine Tech. Rep. No. DOT/FAA/AM-96/26). Washington, DC: Federal Aviation Administration.

Prinzo, O.V. & Britton, T.W. (1993). *ATC/pilot voice conversation: A survey of the literature* (DOT/FAA/AM-93/20). Retrieved from http://www.faa.gov/data_research/research/med_humanfacs/oamtechreports/1990s/media/AM93-20.pdf

Prinzo, V. O., & Morrow, D. G. (2002). Improving pilot/ATC conversation in general aviation. *International Journal of Aviation Psychology, 12*, 341–357.

Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, 13* (3), 257-266.

Schuster, D.A. (2013). *The effects of diagnostic aiding on situation awareness under robot unreliability* (Doctoral dissertation, in press). University of Central Florida, Orlando, Florida.

Serfaty, D., Entin, E.E., Volpe, C. (1993). Adaptation to stress in team decision-making. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 1228-1232.

Sexton, J.B. & Helmreich, R.L. (2000). Analyzing cockpit conversation: The links between language, performance, error, and workload. *Journal of Human Performance in Extreme Environments, 5*, 63-68.

Sharit, J. (2006). Chapter 27: Human Error. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics*, (pp. 708-760). Hoboken, NJ: John Wiley Sons, Inc.

Smith, G., Del Sala, S., Logie, R. H., & Maylor, E. A. (2000). Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory, 8*(5), 311-321.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied,* 74(11, Whole No. 498), 1-29.

Spence, C., & Driver J. (1997). Audiovisual links in attention: implications for interface design. In D. Harris (Ed), *Engineering psychology and cognitive ergonomics, Vol. 2,* (pp. 185-192). Hampshire UK: Ashgate Publishing.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. *Psychological Science, 12,* 462–466.

Strayer, D.L, Drews, F.A., & Johnston, W.A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9,* 23-32.

Sternberg, R. J. (2006). *Cognitive Psychology (4th Ed.).* Belmont, CA: Thomson Wadsworth.

Sublette, M., Carswell, C. M., Grant, R., Klein, M., Seales, W. B., & Clarke, D. (2009, October). Anticipated vs. Experienced Workload: How Accurately Can People Predict Task Demand?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pp. 1383-1387. Sage Publications.

Szalma, J. (2009). Incorporating human variation into human factors/ ergonomics research and practice [Special issue]. *Theoretical Issues in Ergonomics, 10,* 377–488.

Trafton, G., & Monk, C. (2007). Task interruptions. *Reviews of Human Factors and Ergonomics, 3,* 111–126.

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior, 3,* 449-459.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual review of psychology, 53*(1), 1-25.

Vincenzi, D., & Mouloua, M. (1998). Monitoring Automation Failures: Effects of Age on Performance and Subjective Workload. In M. Scerbo & M. Mouloua (Eds.), *Automation Technology and Human Performance: Current Research and Future Trends* (pp. 253-257). Mahwah, NJ: Erlbaum.

Watjatrakul, B. (2005). Influences of figure of merit in aircraft design. *The 19th Conference of Mechanical Engineering Network of Thailand*, 1-4.

Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review*, *17*(4), 479-485.

Wickens, C. D. (1976). The effects of divided attention on information processing in tracking. Journal of Experimental Psychology: *Human Perception and Performance*, *2,* 1–13.

Wickens, C.D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and Performance VIII,* (pp. 239–257). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, C.D. (2001). Workload and situation awareness. In P.A. Hancock & P.A. Desmond (Eds.), *Stress, Workload, and Fatigue,* (pp. 443-450). Mahwah, NJ: Lawrence Erlbaum Associates.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*, 159–177.

Wickens, C. D. (2003). Aviation displays. In P. Tsang & M. Vidulich (Eds.), *Principles and Practices of aviation psychology,* (pp. 147–199). Mahwah, NJ: Erlbaum.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50,* 449-455.

Wickens, C.D., & Dixon, S.R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science, 8*, 201-212.

Wickens, C.D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors, 30*, 599-616.

Wickens, C.D., Dixon, S.R., & Seppelt, B. (2005). Auditory preemption versus multiple resources: Who wins in interruption management? *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting,* (pp. 463-467). Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C.D., Goh, J., Helleberg, J., Horrey, W.J., & Talleur, D.A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45,* 360-380.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors, 25,* 227–248.